

Activism that is illegal and also ill judged

Campaigns aimed at environmental protection need to maintain a careful balance between conviction and fresh thinking. Industry and activists should aim for more of the latter.

To act or not to act unlawfully? That is the question that may eventually confront those opposing activities they perceive to be immoral. Expressing opposition and taking action within society's established structures is all very well until one has failed to carry the day. Illegal demonstration can then attract public sympathy and, with history's hindsight, may come to be universally praised. But the ultimate test of convictions is a willingness to suffer the consequences — possibly martyrdom.

That is the high-flown justification of unlawful activism. But cool-headed judgement as well as passionate conviction is advisable before setting out on that route. Protesters need to demonstrate good sense in keeping illegal actions as a last reluctant resort and matching the damage thereby done to the strength of the cause. Their targets, meanwhile, need to judge well where to give ground. Recent events in the oil and agricultural industries suggest that judgement is questionable on both sides.

Greenpeace is claiming victory last week in a skirmish with British Petroleum — but the claim is dubious, and the skirmish was questionable at its outset. Greenpeace's concerns are fair enough. It wants to do what it can to ensure that environmental impacts of oil exploration are curtailed and carbon dioxide emissions minimized. As BP's chief executive acknowledges, the use of petroleum is responsible for 40 per cent of mankind's carbon dioxide emissions, BP's products contributing 1 per cent.

With profits of over £2 billion (US\$3.2 billion) and rising, one might conclude that BP can withstand a little harassment. Yet its track record puts it at the saintlier end of the industry — it has shown strong support for photovoltaic energy research, a new commitment from the top to establishing greater consensus in responding to the environmental problems associated with the industry, and a distanc-

ing from the anti-regulation lobbyists, the Global Climate Coalition. Here is a significant shift, and yet, to judge by Greenpeace's obstruction of BP's activities in the North Sea (see page 816), the environmentalist group is refusing to recognise that change, effectively resorting to illegal activities as a matter of routine. BP was not only within its rights but well advised to threaten to sue Greenpeace UK for sums that might well have extinguished the organization. Greenpeace's mind needs to be concentrated on opportunities for constructive engagement — and martyrdom is for individuals, not organizations.

Greenpeace's activities — partly but only modestly illegal — in its campaign against genetically engineered seeds show better judgement. Calling attention, for example, to dubious planting experiments (see *Nature* 382, 746; 1996), dumping possibly transgenic soya beans outside a biotechnology meeting — all are fair enough if the intention is to draw public attention.

A more thoughtful approach is that adopted by some US scientists to extensive but, in their view, ill-advised planting of transgenic crops (see page 817). Here the US Environmental Protection Agency is considered by critics to have granted a licence without due caution. Industry (in particular Monsanto and Novartis Seeds) is proceeding legally but in a manner that carries consequences for the environment, in the shape of a possible increased pathogenic resistance to antibiotics. But the activists, rather than acting illegally, are drawing up an alternative research agenda for presentation to the regulatory body concerned.

While Greenpeace should be rethinking, industry too has something to learn. One company is reportedly withholding transgenic seeds from critics wishing to test for resistance frequency. That is another kind of poor judgement. As the oil industry can testify, what better way to stimulate illegal activism and public hostility? □

Science loses in new ministry

Plans to merge Japan's Science and Technology Agency with the Ministry of Education need more thought.

There is good and bad news for science in Japan's plans to streamline its government ministries and agencies (see page 815). The good news is that the Council for Science and Technology, the highest science-policy making body in the government, chaired by the prime minister, is likely to get some teeth. It will have its own secretariat and possibly its own offices, instead of being dependent on the Science and Technology Agency (STA) — which is among the bodies it is supposed to oversee. Whether it will be able to coordinate the activities of the revamped ministries and ensure that each is making effective use of taxpayers' money remains to be seen.

The bad news is the plan to merge the STA with the Ministry of Education, Science, Sports and Culture (Monbusho) and to scrap the idea of a new ministry of science and technology. The prime concern of Monbusho has always been education, not research in the university system. The danger of such a merger is that science and technolo-

gy will continue to take second place to education in the new ministry. Japan's government administration of science and technology looks set to be weakened at a time when it needs to be reinforced to meet the challenges of the next century.

Monbusho bureaucrats greatly outnumber those of the STA. Given that research and development and safety aspects of nuclear power at the STA may be shifted elsewhere, with the associated staff, Monbusho's dominance looks set to grow even further.

The idea to merge the STA and Monbusho came from Prime Minister Ryutaro Hashimoto, who is understandably frustrated with the agency's poor handling of a series of scandals at the Power Reactor and Nuclear Fuel Corporation (see *Nature* 386, 746; 1997). But to decide the future administration of Japan's science and technology on the basis of some recent politically-sensitive bungling seems short-sighted to say the least. □



Radical restructuring in Japan merges science and education

[TOKYO] Hopes that an independent ministry for science and technology might emerge from Japan's continuing administrative reform appear to have been dashed.

According to new plans for the restructuring of ministries released last week, Japan will instead create a single Ministry for Culture, Education, Science and Technology from a merger of the present Ministry of Education, Science, Sports and Culture, which oversees research in universities, and the Science and Technology Agency (STA).

The plans have been put forward by a government panel, headed by the prime minister, Ryutaro Hashimoto, responsible for formulating a restructure of ministries planned for 2001. They will be included in an interim report due for publication later this year. A parliamentary decision on the reform plan is expected by next spring.

While rejecting the concept of a ministry of science and technology, intended to bring together existing mission-oriented research programmes and national research centres, the panel recommends promoting Japan's relatively powerless Environment Agency to ministry rank. In an earlier proposal the agency would have been merged with the STA into a Ministry for Science, Technology and the Environment.

Akito Arima, president of the Institute for Physical and Chemical Research, represented the interests of the scientific community

on Hashimoto's panel. Science policy was a rather "marginal issue" in last week's deliberations, he says, and many questions have been left open.

The proposed merger raises several issues, in particular the fate of the mission-oriented research bodies attached to STA — including the much criticized Power Reactor and Nuclear Fuel Corporation and Japan's National Space Agency. Many feel that these institutions would not fit happily into the education ministry. It will be a difficult task to bring together the "top-down oriented STA with the slow moving, bottom-up oriented ministry of education," says Arima.

Some say that integrating the STA into the ministry of education is the worst scenario possible, because the education ministry's preoccupation with school education could overshadow science and technology. But others see benefits for the scientific community.

Tsutomu Kimura, president of the Tokyo Institute of Technology, says that the merger should make science and technology policy in Japan "more straightforward, efficient and effective". It could resolve some absurd situations that arise because of the bureaucratic wall between the ministry of education and the STA — such as the fact that university researchers receiving large STA grants often have to set up second laboratories outside the university campus because of educa-



tion ministry regulations.

Kimura also argues that integration with STA may promote the adoption of objective research assessment and evaluation in Japanese universities, which the education ministry has not pushed forward.

Some senior STA officials regret that the agency was not included in the political decision-making which has arranged their fate. They say that this exclusion will be damaging to the morale of the 10,000 staff in STA-related research. They say that there are still many details to be sorted out in the proposed reform.

"We have just seen the birth of the basic law on science and technology last year," comments Kaname Ikeda, director STA's nuclear safety bureau. It will be important to watch how the proposed ministry promotes the "spirit" of that law, Ikeda says.

But even within STA there are some who see a positive side to the proposed merger with the education ministry. Breaking down the barriers between national laboratories and universities could help to improve scientific and technical education, they say.

Other details of the restructuring, as well as the fate of national research institutes attached to other ministries — including the laboratories of the Ministry of International Trade and Industry's (MITI) Agency of Industrial Science and Technology — are expected to be discussed at future meetings of the reform panel.

A sensitive issue concerns the administration of Japan's nuclear programme, currently divided between the STA and MITI. Nuclear energy research will probably be handed over to a future Ministry of Industry — a kind of MITI stripped of its industrial policy ambitions. Nuclear safety may be delegated to the proposed environment ministry.

Richard Nathan & Robert Triendl

New roles for science and technology council

[TOKYO] Now that proposals for a Japanese science and technology ministry have been dropped, one expectation is that the Council for Science and Technology will take over some of the tasks and functions of the present Science and Technology Agency, possibly including the administration of big-science programmes. The council is Japan's highest science policy-making body and is chaired by the prime minister.

At present, the STA oversees an important part of Japan's big-science programmes. As most of these — including nuclear fusion and space — have a heavy tilt towards

development, many observers wonder how they could fit into the framework of the proposed new education, science and technology ministry. But others say that the proposed reform opens up an opportunity to render existing projects at STA both more cost-efficient and more scientifically rewarding.

In the future, the social sciences and humanities will be added to the council's constituency — which at present covers only natural sciences and engineering. And the council will enjoy a considerably enlarged budget, with more authority and staff of its own, to help it fulfil its originally intended role of coordinating and

overseeing all government science and technology.

"The fact that the council does not have administrative staff, but relies almost entirely on STA personnel, impinges on its ability to act as a coordinating body," says Wataru Mori, a permanent member of the council.

Akito Arima of the Institute for Physical and Chemical Research hopes that the council can play a more active role in Japanese science policy. At present, the council's main activity is to report on policy questions handed down to it by the prime minister. "A future Council for Science and Technology should play a more active role in policy-making," he says. **R.N. & R.T.**

Greenpeace defiant in face of threat from oil company

[LONDON] The environmentalist group Greenpeace says it has no intention of scaling down its disruption of oil exploration, despite being threatened with bankruptcy by the oil giant British Petroleum (BP).

A court in Edinburgh, Scotland, last week froze Greenpeace's UK bank accounts, issued a writ for £1.4 million (US\$2.17 million), and indicted four activists for losses incurred from delays caused by Greenpeace's week-long occupation of an oil platform earlier this month. The activists had chained themselves to a mobile drilling platform which was being towed to the Foinaven oilfield. The oilfield is part of new reserves west of the Shetland islands, known as the Atlantic Frontier.

BP withdrew its damages claim after two days, as well as charges against three of the activists, in return for an undertaking that Greenpeace would stay away from Foinaven. The company had earlier tried to obtain an assurance that Greenpeace would not obstruct BP activities in the whole of the Atlantic Frontier. But Greenpeace said it would not give such an undertaking.

Marcus Rand, a Greenpeace campaigner, says the pressure group may continue direct action in the Atlantic Frontier's other main oilfield, Schiehallion. He says Greenpeace will try to prevent all new oil exploration on the grounds that increased fossil fuel use will contribute to climate warming.

Rand says BP's withdrawal is another triumph for the environmentalist group. But some observers believe that the oil company, and not Greenpeace, emerged as a narrow victor in this particular skirmish.

One oil industry source says BP probably

never intended to carry out its threat of extracting damages, but sought to publicize the fact that Greenpeace was breaking the law. He says Greenpeace will have been wounded by being compelled to abandon an important weapon in its campaign arsenal: taking direct action.

But Rand says: "Direct action is a core value of Greenpeace, and will remain a core value. We have had injunctions served on us in the past, but they have not prevented us from taking direct action."

BP's court move could mark a shift in the oil industry's tactics in its battles with the environmentalist movement. It also suggests the industry learned valuable lessons in the wake of Shell's defeat at the hands of Greenpeace two years ago over its plans to sink the Brent Spar oil platform in the North Atlantic (see *Nature* 376, 378; 1995).

Greenpeace mobilized an international boycott of Shell petrol (gasoline) stations in response to Shell's insistence on deep-sea disposal of the oil platform. This time, BP gave Greenpeace no chance to organize another international campaign when it cancelled the writ for damages two days after it was issued.

Unusually for an oil company, BP has consistently sought to raise its public profile as an environmentally conscious enterprise. In May, it announced it would no longer oppose the scientific consensus that human activity was contributing to climate change — the first oil company to do so.

It also announced a significant increase in its investment in solar energy research and pulled out of the Global Climate Coalition, an association of oil and automobile companies that has voiced doubts about the scientific consensus on global warming.

But Greenpeace distrusts BP's environmentalist credentials, particularly because BP is one of 22 oil industry defendants in a separate action brought by Greenpeace, which asks the courts to investigate the legality of oil exploration licences given by the UK government on the Atlantic Frontier. Greenpeace claims that the UK government is in breach of European law as it did not conduct an environmental impact assessment before granting the licences.

BP has conducted its own private environmental impact assessment, which, a spokesman says, was the largest and most comprehensive of its kind. Sarah North, a Greenpeace marine campaigner, agrees, but says the study raises more questions than it answers. "There is no real analysis of the impact of spills on the coast, nor any specific data on the benthic organisms in the immediate vicinity of Foinaven." **Ehsan Masood**

Karolinska Institute disowns work of cancer researcher

[PARIS] The Karolinska Institute in Stockholm, Sweden, is distancing itself from the work of one of its former researchers, Ulf Lönn, six months after an independent panel of investigation concluded that he had "most probably" committed scientific fraud.

Lönn denies the charges. But, according to the investigation, about 20 papers published in the past decade by Lönn, who works on gene amplification in breast and bladder cancer, included data that appeared to have been manipulated. "A large number of Lönn's publications must be recalled and disclaimed as unreliable," says the report.

The Karolinska Institute is now considering introducing a compulsory system of archiving researchers' raw data with the aim of making manipulation more difficult.

The institute stops short of accusing Lönn formally of misconduct. Hans Wigzell, the vice-chancellor, says that from a legal point of view "we cannot conclude beyond all reasonable doubt that the data could not have come about by means other than fraud." Nonetheless, Wigzell says, the weight of evidence against the researcher is such that the institute has decided "not to stand by this research, because we consider it is irreproducible science". He will send letters to funding agencies and scientific journals this week explaining the conclusions of the investigation.

Suspicions were first raised early last year by one of Lönn's colleagues, who felt that the stream of publications emerging from his group exceeded what would be expected from the level of laboratory activity. Shortly afterwards, the institute set up a three-man panel of outside experts, chaired by Jan Pontén, professor of pathology at the University of Uppsala, to investigate the allegations. The panel delivered its report in February.

Lönn believes that he has been legally "cleared". His lawyer, Frederick Bostrom, from the Stockholm-based firm Lindwall and Nordbäge, says that the investigation did not give the scientist "a fair chance" to defend himself, and he may bring a court case.

According to one scientist involved in the investigation, the Karolinska Institute took six months to react to the panel's finding because it was overly concerned about its image. The institute awards Nobel prizes in medicine and physiology. Pontén says that the institute should have alerted scientific journals to the allegations earlier.

Wigzell says the possibility of taking further action was complicated because the institute did not directly employ Lönn, who was funded by the health authorities. Lönn has recently left the institute. **Declan Butler**



Chain gang: Greenpeace activists during their week-long occupation of the oil platform.

Dispute over insect resistance to crops

[WASHINGTON] Academic scientists and industry are in dispute about efforts to prevent the emergence of insect resistance to new, genetically engineered crops, a development that would render useless a naturally occurring insecticide relied on by organic farmers.

At issue are corn and cotton engineered to produce a toxin made by the soil bacterium *Bacillus thuringiensis* (*Bt*). The toxin defends plants against common pests including cotton bollworm and tobacco budworm, which both afflict cotton, and the European corn borer. *Bt* cotton has been developed by Monsanto; the major manufacturer of *Bt* corn is Novartis Seeds. Crops were first planted commercially in 1996.

About 2.1 million acres, or 15 per cent of the US cotton crop, have been planted with *Bt* cotton this season. *Bt* corn has been planted on about 7.25 million acres, or 9 per cent of corn acreage.

Critics, led by the Union of Concerned Scientists (UCS), argue that large-scale planting will put such intense genetic pressure on insects that resistance to *Bt* toxin will emerge rapidly. They claim resistance management plans (RMPs) are inadequate.

The Environmental Protection Agency (EPA) worked out provisional RMPs with companies as part of conditional approvals for planting *Bt* crops. Further development of these is required if approval is to be extended beyond 2001.

Current resistance management strategies for *Bt* allow resistant insects that survive in *Bt*



Cotton harvest: infestations of the genetically engineered crop aroused fears about resistance.

fields to mate with non-resistant insects from adjacent "refuges" where non-*Bt* plants are grown. The aim is to drown resistance genes in a larger gene pool.

But the size of refuge necessary to ensure resistance genes are drowned out is disputed. Under Monsanto's plan, farmers must plant at least four per cent of their cotton crop with non-*Bt* plants. Bruce Tabashnik, head of the entomology department at the University of Arizona, one of UCS's scientific advisers, says that 20–40 per cent would be safer.

Fears were prompted by infestations of Monsanto's *Bt* cotton in Texas last year by cotton bollworms (Nature 382, 289; 1996). The company and the EPA say that unusually large populations of bollworm were responsible, but critics claim that levels of *Bt* generated by the cotton plants were not high enough. They say that if the cotton is to continue in use,

refuges should be increased to ensure they hold enough non-resistant insects.

Monsanto spokesman Gary Barton says that the plan was developed using the "best available science" and points out there is an option for a 20 per cent refuge which growers may spray with non-*Bt* insecticides to protect crops without fostering *Bt* resistance.

Critics say too little is known about the issues — from insect movement and mating patterns to naturally occurring rates of insect resistance — to make current RMPs work effectively, and that companies are not doing the research that would improve them.

To bolster its case, the UCS will summon six scientists to Washington next month to draft alternatives to existing RMPs. Environmentalists are also agitating for the EPA to seek outside advice on the soundness of its resistance management science, a move the EPA is considering.

Tabashnik says that current EPA-approved plans are "based on economic acceptability to the companies selling the products".

But Janet Andersen, director of EPA's biopesticides and pollution prevention division, says it is "premature" to judge the effectiveness of the plans because crops are in only their second commercial season.

Jeffrey Stein, regulatory manager at Novartis, calls it "ludicrous" to suggest that companies would neglect RMPs in the quest for short-term profits. He says Novartis will spend more than \$250,000 on resistance research this year because, while resistance of insects in the wild to *Bt* corn is "strictly hypothetical" at the moment, "it may happen".

Monsanto's Barton agrees and says that the benefits of the new crops far outweigh the risks, increasing yields while eliminating the use of one million litres of chemical insecticides. "Do the critics just want to keep spraying?" he says.

Refuges for corn crops are not required under current RMPs, a further cause for criticism. Industry argues that, with 91 per cent of US corn acreage not planted with *Bt* corn, ample natural refuge already exists for present needs.

Meredith Wadman

Swiss research resting on its laurels

[MUNICH] Swiss science is still of high quality, but it is living off its past successes, says a report from the Swiss Science Council that analyses the state of research and universities.

The report warns that recent years of falling investment in research mean current high standards "cannot be guaranteed in the future".

Science council reports are prepared every four years and are used by the federal government to help develop its short-term science and higher education policies. The current report has been prepared for the 2000–2003 financial planning period.

Since 1992, Swiss spending on research has stagnated at around Sfr9 billion (\$6 billion), and the proportion of this provided by industry has fallen from 75 per cent to less than 70 per cent. The report does not criticize the relative reduction in industrial research and development investment, since the proportion of spending on research by industry in Switzerland has always been much higher than in other European

countries. But it says that the state should make up for the reduction in future.

The report says new money should be used to support Switzerland's strengths such as biomedical and clinical research, computational physics and material sciences.

A second report on universities says that support for the so-called *Mittelbau* in Switzerland — the 20,000 or so doctoral students, research fellows, assistant and associate professors — is a "central problem of the science system". The *Mittelbau* are usually on short-term contracts and have little responsibility within their institutes, whose hierarchical systems deny them freedom in their research and freedom to teach.

The report says working conditions and career prospects of the *Mittelbau* should be improved by the universities and the federal and canton governments. Training and support programmes should be available and grants should be provided to encourage the notoriously non-mobile Swiss PhD students to study abroad. **Quirin Schiermeier**

Mir's problems could mean scale-down for science

[WASHINGTON] The apparently successful restoration of power to the Mir space station last week (the power had not been fully restored as *Nature* went to press) means that onboard research should resume when a new American astronaut arrives on the station next month. But many experiments will be scaled down from their original design, and the long-term utility of Mir as a research platform is still in question.

Mission managers at the US National Aeronautics and Space Administration (NASA) say that 80 per cent of the original science objectives for the US/Russian research programme on Mir can still be met, even with the damaged Spektr module out of commission and power levels between half and three-quarters of what they were before a cargo ship crashed into Mir last June, damaging its solar arrays.

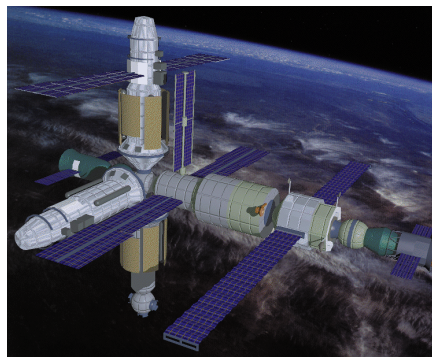
But John Charles of the National Aeronautics and Space Administration (NASA) Johnson Space Center, a mission scientist for the joint research programme, says that estimate is "a little bit generous".

In a revised research programme approved by both Russian and US investigators, some experimenters will have data collected less often to conserve power, while others will carry out only a subset of their planned investigations. Richard Zwierko, a shuttle-Mir mission planner at NASA headquarters in Washington DC, says the guiding philosophy behind the revised programme is that scientists would get the "minimum data" they need to obtain a useful result.

But several factors may further reduce the science output. The Kristall and Priroda laboratory modules, which have been shut down since June, may not be in full working order even if power is restored. The Mir crew has done its best to mop up condensation that formed inside the modules, but no one is certain if the moisture has damaged any equipment. Kristall and Priroda are used for microgravity and remote-sensing experiments.

Manpower is another concern. Dave Wolf, the US astronaut who will operate most of the experiments on Mir starting in September, will also be helping the Russian cosmonauts to repair the station. Wolf was a last-minute replacement for Wendy Lawrence; the switch was made when NASA decided to send an astronaut who could use the Russian space suit (Lawrence is too small). Wolf will take at least one spacewalk, which had not been planned for Lawrence's mission.

Wolf is not as familiar with the Mir experiments as Lawrence was, and is training furiously to catch up before his planned launch on 25 September. Biomedical researchers are



Science is a low priority on board Mir.

struggling to collect baseline data on the astronaut before he launches; this is critical to evaluating the physiological changes that occur in space. Without such data, some in-flight experiments would be degraded.

Although Wolf is doing his best to accommodate the data collection, according to Charles, he has many competing demands on his time, including training for his spacewalk. As a result, says Charles, "there may still be holes" in the baseline data when he launches.

Based on recent experience on board Mir, Wolf can also expect regular computer failures and other hardware glitches on the ageing space station to require him to drop experiments and pick up a wrench. One non-NASA scientist participating in the shuttle-Mir programme says that, although "the astronauts do the best they can, science experiments are, for better or worse, at the bottom of the totem pole".

Charles says this has always been the case. Science, he says, was originally last on NASA's list of justifications for participating in the shuttle-Mir programme, after considerations such as gaining experience cooperating with the Russians and practising docking techniques. Indeed, NASA now uses "operational experience" as its primary rationale for continuing use of Mir — repair work and troubleshooting are seen as valuable in themselves, as they are similar to the kinds of maintenance that will be done on the International Space Station, construction of which is scheduled to begin next summer.

Russia, meanwhile, is reluctant to abandon Mir as long as its space engineers are persuaded that each new problem can be fixed. And the United States is loath to pull out unilaterally. So, barring a life-threatening emergency or major system failure, Mir is likely to keep flying, even with reduced capability.

According to Zwierko, even if onboard power never again gets above the 50 per cent level, some of the research programme will still be possible.

Tony Reichhardt

State seeks independence in science policy

[CANBERRA] Victoria has become the first of Australia's six states to develop a science and technology strategy that is independent of federal government policy.

The state premier, Jeff Kennett, has established a 19-member science, engineering and technology taskforce which he will co-chair with Adrienne Clarke, a plant cell biologist at the University of Melbourne. Clarke is a former chairman of the Commonwealth Scientific and Industrial Research Organization, and was recently made Lieutenant Governor of Victoria with a unique brief to promote science in the state (see *Nature* 387, 447; 1997).

As part of the drive, Victoria has also created the post of Principal Advisor for Science and Engineering, which has been taken by Margaret Britz, a food microbiologist from the Victoria University of Technology.

The commitment to developing an independent science policy contrasts with the approach of the federal government which has been criticized for indecisiveness over research during its 18 months in office.

Victoria is Australia's second most populous state and the most industrialized, with A\$33 billion (US\$24 billion) — or 30 per cent of its gross product — being generated from science and technology. Mark Birrell, the Victorian minister responsible for science, says his prime aim is to improve the business environment for research and development, a popular move among industry leaders who are openly frustrated by adverse actions of the federal government.

Although some new projects have been announced at modest levels of funding, the key to Victoria's future scientific success will be its capacity to finance substantial long-term programmes. All states have long felt frustrated by their dependence on taxes collected federally. Victoria has gone further than others in breaking away from subordination to Canberra by generating significant revenues from gambling in a vast new casino and from privatization of public utilities. Both of these moves are controversially promoted by Kennett.

In a statement launching the initiative, Kennett said its success "will depend on progressive Commonwealth reform of funding in areas such as taxation, research and development funding opportunities and intellectual property law". This is seen as a criticism of the federal government's recent science funding cuts.

Peter Pockley

NAS president hints he may run again

[WASHINGTON] Bruce Alberts, president of the US National Academy of Sciences (NAS), will decide next month whether to put himself forward for a second six-year term as head of the leading US scientific body.

In a move that hints that he is preparing to stay on after his first term expires at the end of 1998, Alberts last week met Richard Klausner, director of the National Cancer Institute (NCI), to discuss how he could renew his involvement with cancer research through the NCI, which is based just outside Washington in Bethesda, Maryland.

Alberts said in an interview that he was shifting the emphasis of his leadership from science education to the promotion of international science to support sustainable development, especially in poor countries.

He added that he wanted to stay in the job, but would balance that against his desire to do research and return to his family home in California. After a review of his performance last month, Alberts said, the academy's ruling council "told me they'd like me to run for a second term, while recognising that there are personal reasons why I might not want to do so. I told them my wife and I would have a vacation in early September and come back and give them an answer."

The council will set up a committee in the autumn to decide who should be put forward for election to the presidency: in previous elections, only one name has been on the ballot sent to the academy's 1,800 members.

Alberts says that he would not be deterred from running by the NAS's continuing struggle to reverse a January court ruling that it must reorganize committee meetings of its operating arm, the National Research Council, to comply with the Federal Advisory Committee Act (FACA). The academy says the



change would undermine its independence from government (see *Nature* 386, 309; 1997).

"I like challenges," he says. The FACA struggle is "something that makes it more likely that I will run again." The January ruling, he says, "is so illogical that I can't imagine that it won't be fixed in the end."

Jack Gibbons, President Clinton's science adviser, said on 23 July that FACA was "a thing that you have to learn to live with, and the academy may need to live with it".

But Alberts says he has no regrets about the way in which the FACA issue has been handled, and dismisses suggestions that, instead of highlighting its damaging effects and filing an appeal to the Supreme Court, as it did last month, the NAS should have moved quietly to work within FACA's constraints.

The FACA ruling "is the most serious crisis the academy has faced in the last 10 years at least," says Alberts. It "would really mean that the independent voice of the academy would be lost."

Although the Department of Energy has halted several of its studies as a result of the

FACA ruling, other departments have continued business as usual and the NRC budget — chiefly derived from government contracts — is being maintained.

"Our projected budget is the same as last year, and that surprises us because we were fearful that we'd have a major drop," says Alberts. A pay freeze imposed earlier in the year (*Nature* 387, 220; 1997) has been lifted and staff given an average 2 per cent increase, although senior officers' pay stays frozen. "It was my decision to take a hard line on salaries, to make sure we were protected."

Alberts acknowledges criticism of his reluctance to lobby directly for science funding, but he has no plans to "behave the same way as all of the scientific societies" by doing so. "We've been taking a broader view of how we support science than the view I think some people would like us to take," he says. But he identifies funding at the US Department of Agriculture (USDA) as a concern, and will continue to press for a sharp increase in USDA's peer-reviewed grant programme, the National Research Initiative, from \$100 million to \$500 million a year.

He also pledges to try to persuade the State Department, which has recently downgraded its scientific operations, to recognize the importance of science in international relations. He describes Tom Pickering, former US ambassador to Russia and now deputy secretary of state, as "a close friend of science".

Having started his first term in 1993 with an emphasis on the promotion of science education for the public, Alberts now says another priority — to support "sustainable development" around the world through science disseminated via the Internet — has equal billing. He says better communications mean that "we can create a completely different science" for dealing with issues such as clean water supplies. "I want the academy to take a lead on this."

These two priorities, together with the promotion of "excellent and vigorous" science to support them both, are the three items Alberts has selected for attention in a second term, if he serves one. "Our job is to do these three over the long term, and not to get distracted by short-term issues."

This means he has no plans for bold reform in the structure of the academy itself, though issues of openness and accountability have been raised by the FACA dispute.

He says: "The main issue is not the 1,800 members of the academy. It is the million scientists in the United States, where they are and what their role is in society. That is where I will focus, not on details I can't affect and which don't really matter."

Colin Macilwain

Cooperative Centres fight cuts proposal

[CANBERRA] Tony Staley, president of Australia's governing Liberal party, has challenged a government report on industrial policy that recommends cutting funding for the 65 Cooperative Research Centres. The CRC scheme forges partnerships between universities, government agencies and industry for targeted research and development.

In a letter last week to John Moore, the Minister for Industry, Science and Tourism, who commissioned the report from David Mortimer, a businessman, Staley said the cuts would "destroy the CRC programme".

The government is also investigating making CRCs "more self-funding". At present they receive an annual government budget of A\$146 million (US\$108 million), which Mortimer says should be reduced to A\$20 million and used only for "public good" research.

Staley accuses Mortimer of "lack of understanding of the CRC programme". This programme has been successful, with much of the research directed to industrial innovation, he says. Influential businessman Sir Arvi Parbo also criticized the report in a letter to the Australian Academy of Technological Sciences.

Australia's 36 universities hit out at Mortimer's view that they should depend more on funds from industry. The Australian Vice-Chancellors' Committee demanded the government "reject totally Mortimer's recommendations concerning R&D".

The committee says "swingeing cuts" to research in universities, CRCs and the Commonwealth Scientific and Industrial Research Organisation would make Australia's economic growth "critically dependent on research owned offshore".

Peter Pockley

Universities reform makes headway in Germany

[MUNICH] After months of acrimonious debate, Germany's federal and state research ministers last week finally agreed an amendment to the federal framework law on universities, the *Hochschulrahmengesetz*, which would give Germany's 325 universities much more room for manoeuvre.

The proposed amendment would give universities more financial and administrative autonomy and would reduce the number of rules that the federal law imposes on them. Many of these rules had been intended to guarantee a countrywide uniformity in higher education standards. But in recent years they have been held responsible for the inefficiency, bureaucracy and low international competitiveness prevalent in the universities (see *Nature* 384, 204; 1996).

If the amendment is approved by parliament in the autumn, universities would gain significant new rights. They would be able to choose, at least in part, their students — at the moment students are imposed on universities through a central organization. They would also be able to offer internationally recognized bachelors and masters degrees.

University academics would also have to compete more for public funding. The quality of research and the commitment to the education of young scientists would be major funding criteria. All faculties would therefore have to undergo regular external evaluation, with transparency being the name of the game. Students would vote on their professors' teaching performance, and the results would be published. A large lobby of university professors has opposed moves to change the status quo.

Oxford scientists attacked by animal activists

[LONDON] **Property belonging to five scientists at the University of Oxford, including Colin Blakemore, professor of physiology, was attacked last week by animal rights activists. The Animal Liberation Front (ALF) has admitted to carrying out the attacks to draw attention to a jailed member who is on hunger strike.**

Barry Horne, an inmate of Bristol prison, is protesting at the government's refusal to convene a royal commission on the use of animals in research, despite a pre-election promise to do so (see *Nature* 388, 311–312; 1997). The scientists' cars were damaged and their houses sprayed with graffiti. An ALF spokesman said "the severity of actions can be escalated".

'Higgs-boson' Higgs wins again

[LONDON] Peter Higgs, emeritus professor of physics at the University of Edinburgh, but better known for the particle named after him, the 'Higgs boson', has been awarded the European Physical Society's High Energy and Particle Physics Prize for 1997.

Higgs will share the prize with the Belgian physicists Robert Brout and François Englert. The three scientists are being honoured for their work in unifying the electromagnetic and weak forces.

This is the third time Higgs has been honoured this year. He received the 1997 Paul Dirac Medal from the UK Institute of Physics, and in May was awarded an honorary degree from the University of Bristol.

Scientists defend stand on fish stocks

[MONTREAL] **Three Canadian government scientists have hit back at charges made earlier this year in an article in the *Canadian Journal of Fisheries and Aquatic Sciences* that political and bureaucratic interference in fisheries science had compromised the government's efforts to sustain stocks of Atlantic cod (see *Nature* 388, 106; 1997).**

William Doubleday, director-general for

science in the Department of Fisheries and Oceans (DFO), and two senior officials for the Newfoundland region say the fisheries department invited independent examination of the cod stock and had followed its advice.

"Rather than suppressing information about the importance of high fishing mortalities in the decline of Atlantic groundfish, DFO stock status reports have highlighted its importance," the officials say. "We consider that the peer-review process in place for Canadian fish stock assessments is as good or better than other systems in use elsewhere in the world."

California checks its minorities

[SAN FRANCISCO] The US Labor Department says it will investigate graduate admissions at the University of California (UC) to determine whether abandoning affirmative action policies has resulted in employment discrimination. Graduate students are the labour pool for many jobs on campus, including research assistants.

In 1995, after a California voter initiative banned race-based or gender-based preferences in hiring, contracting and education by state institutions, UC halted race-conscious admissions policies intended to integrate its campuses. The measure is

under court review, but the school has already dismantled the policies.

The university has nine campuses, manages three federal laboratories and is the largest US government contractor. Under federal law, government contractors must make special efforts to improve hiring of women and minorities.

First patent on breast cancer gene

[WASHINGTON] A US biotechnology company has been granted a US patent on the coding sequence of the BRCA1 gene, which is associated with hereditary breast and ovarian cancer. OncorMed of Gaithersburg, Maryland, was issued with the first US patent protecting a full-length coding sequence of the gene on 5 August. OncorMed may file for similar patents on the gene sequence in other countries.

The company says that the patented sequence represents the healthy form of the gene in more than 50 per cent of the population, making it valuable in identifying individuals not at risk of inherited cancer, in finding previously unknown mutations, classifying tumours, and developing anticancer drugs. A statement said that the patent "reflects OncorMed's intention to aggressively protect its BRCA1 gene-related intellectual property worldwide".

New chief for British Geological Survey

[LONDON] David Falvey, director of the international Ocean Drilling Program (ODP) at the Joint Oceanographic Institutions in Washington DC, has been appointed as the new director of the British Geological Survey. Australian-born Falvey is an exploration geophysicist. He worked with the Shell oil company and the University of Sydney, and was an associate director of the Australian Geological Survey Organization before joining the ODP.

Falvey succeeds Peter Cook, director of the British Geological Survey since 1990. Cook is widely credited with successfully steering the organization through one of its toughest periods.

Munich office of *Nature* on the move

The Munich office of *Nature* will be moving to the following address from 25 August:

● Josephspitalstrasse 15, D-80331 München
Telephone (editorial) +49 89 549057 13
Telephone (classified advertising) +49 89 549057 12,
Fax +49 89 549057 20,
e-mail address remains as before:
a.abbott@nature.com

Peer review is a two-way process

Sir — We too are concerned about the fairness of the grant-awarding process. So far, recent interest has come from the applicant's viewpoint (Wennerås and Wold, *Nature* **387**, 341–343 & L. M. Castell, **387**, 841; 1997).

We write from the other side. Not every applicant, referee or grant committee member is free from prejudice or vested interest and this makes the task of managing the grant-awarding process especially challenging.

We have identified a number of problems and made some changes to our procedures:

- The chairperson, to avoid conflicting loyalties, does not apply for grants, in accordance with Association of Medical Research Charities guidelines for peer review (1995). Failure to observe this commonsense rule effectively renders any grant-awarding committee leaderless and vulnerable to self-interest (A. Fielder, *Lancet* **347**, 1188; 1996).

- Referees' comments are fed back to the applicant for brief comment before the meeting, thus involving the applicant prospectively, as in R. P. Clarke's suggestion for paper refereeing (*Nature* **386**, 319; 1997). We believe this is critical: after all, the applicant may well be the expert and best able to spot factual errors and prioritize criticisms. This helps the committee, which frequently lacks specific expertise, and reduces the to-ing and fro-ing which can retard the decision-making process. We hoped that this approach would encourage referees to provide more balanced criticism, but unfortunately it did not. Indeed, selecting the appropriate referee for each topic and ensuring evenness across the research field remains a major weak link in the process.

- We have now broken with tradition and a committee member who is also an applicant is no longer permitted to defend his/her application in person, so reducing any

advantage over an applicant who is not on the committee.

One must also beware of the scientist who declares that only "good science" should be supported. This phrase hides a multitude of sins, of which self-interest is a consistent feature, and is an effective method of eliminating competition. Memories can be short — after all, early in our careers many of us had lucky breaks, some of which had their origins in serendipity rather than scientific merit. New talent and innovation must be supported and sometimes this means taking risks.

Many problems remain, linked perhaps by the common thread of a lack of transparency and secrecy in the grant-awarding process. This really is not necessary — it is understood that tough decisions have to be made and not all applications can be funded. We must do better. Far too much time and energy is being wasted on a process that is stacked in favour of funding "more of the same" rather than innovation. So, let's keep the debate going to devise more efficient and equitable ways of funding science.

Alistair Fielder
Hannah Vinyard

*Prevention of Blindness Sub-Committee,
Royal National Institute for the Blind,
224 Great Portland Street,
London W1N 6AA, UK*

Sir — Every investigator will empathize with Linda Castell's frustrations with the peer-review process. However, her reference to the article by D.F. Horrobin (*Lancet* **348**, 1293–1295; 1996) as a possible alternative is hopefully more in jest than conviction.

Horrobin proposed abolishing the UK Medical Research Council and persuading research charities to abandon peer review. That would allow upwards of £100,000 to be given each year to every academic clinician in the United Kingdom to spend on whatever ideas come into his or her head.

Peer review is one part of a complex dynamic that is continually sapping the energies of funding agencies and grant seekers alike. The UK biomedical research infrastructure is under continuous strain from research assessment exercises, year-on-year reductions in government funding, policy changes by research councils, disequilibrium between research councils and charitable funding, declining infrastructure and so forth.

All these pressures overtly or covertly affect the quality and integrity of the peer review process. Regrettably, funding agencies are all too often seen as passive administrators of their research base, thus compounding the distrust of the peer review process. The response to all this must be for funding agencies to win the confidence of applicants by proactive management of their research portfolio.

The Leukaemia Research Fund debarbs all major grant-holders from sitting on its advisory panel; site-visit committees comprise mainly overseas experts; no group leader can site-visit another group leader; feedback is provided to all applicants (successful and unsuccessful); and resubmissions are invited.

Members of our advisory panel serve for only three years and are drawn from the whole biomedical community. After all, one does not need an expert in leukaemia to assess the quality of a positional cloning project or the role of p53 in drug resistance.

Clearly no system is perfect and we strive continually to improve. Peer review must be a two-way process between applicants and funding agencies. Trust in this dialogue can be achieved only if agencies are open and democratic in the management of their research portfolios.

David Grant

*Leukaemia Research Fund,
43 Great Ormond Street,
London WC1N 3JJ, UK
e-mail: grant@leukaemia.demon.co.uk*

Sacred circle

Sir — I am surprised at the suggestion by Elizabeth Aveling that the Coupland enclosure, Milfield, was a winter kraal for cattle, as that does not fit with the given facts (*Nature* **387**, 553–554; 1997).

Aveling supports Waddington's interpretation that it was used to winter cattle. Yet the fact that a double-ditched linear feature bisects the enclosure indicates that cattle passed through it, not into it; the double-ditched feature is designed to keep them out of the two opposing cords.

The idea of a kraal raises ancillary questions anyway — how many grazing cattle could this relatively small 'field' support, or did the early Neolithic undertake silage production? Is a major earthwork embankment a reasonable way for our ancestors to contain cattle, or is it proposed that the embankment was defence against human marauders?

The real answer, I suggest, lies in the almost throwaway line that it was perhaps a sacred site.

I suggest that the animals were driven

through the henge to be blessed, so that the all-powerful god(s) would protect the herd from pestilence and predators and ensure fertility. The opposing cords of the circle probably contained religious structure, symbols, sacred relics and the tribe priests. The nearby ford may also have played its part in these religious rites.

Roger Matthews

*Drummond Cottage,
Old Tavern Yard,
Westbourne,
Hampshire PO10 8TA, UK*

Space can't wait

Sir — We applaud Alexander Harcourt (*Nature* **387**, 340; 1997) for emphasizing the value of exploring the world around us. Pressure abounds to devote all our resources toward investigations that promise quick technological pay-offs. It is refreshing to read an exhortation for support of discovery and understanding of the biological sphere.

In recent years, our understanding of biota has grown significantly from an unexpected vantage point — the “lifeless void” of space. Only in space can the influence of gravity and other earthly physical phenomena be separated from the organisms and biological processes we want to understand.

By conducting biological research on orbital platforms, scientists have been able to make tremendous gains in basic understanding of human physiology, protein structure, signal transduction, developmental biology and immunology.

Exciting developments in the area of exobiology may signal present or past life on our neighbouring planet Mars, and have also sparked the search for life in extreme environments on Earth. Space research ignites our curiosity and goads us on to new levels of understanding.

We therefore disagree with Harcourt's contention that “space can wait” for the attention of future generations. It is also a mistake to pit one area of science against another in the guise of bolstering funding of earthbound exploration, when contributions to our understanding come from every discipline.

Scientists should instead strive to foster an awareness that scientific inquiry and discovery are intrinsically valuable and worthy of increased support. To borrow a slogan from the International Space Station: “Great nations dare to explore.”

Mary E. Musgrave

(President, American Society for Gravitational and Space Biology)
Department of Plant Pathology & Crop Physiology,
Louisiana State University,
Baton Rouge, Louisiana 70803, USA
e-mail: xp3031a@lsuvm.sncc.lsu.edu

Sir — I was disappointed to read your leading article “Martian life to be avoided” (*Nature* **388**, 211; 1997). I strongly believe that it is appropriate to plan, and to achieve, the human exploration of Mars.

Excellent science is indeed possible without humans, as the robotic explorations of the Moon by the Ranger, Surveyor and Luna probes, and of course Mars by the Viking probes, has proved. However, there is no substitute ‘in the field’

for the flexibility of the human hand and mind; ask Harrison Schmidt, Apollo 17 astronaut and the first geologist to study, first hand, another world.

Apart from the vast increase in our knowledge of our nearest neighbour, we have in fact been bequeathed a huge legacy by the US and Soviet space programmes in the 1960s and 1970s; computer miniaturization, the development of new materials for engineering and improvements in satellites and communications were all accelerated by the space race. And almost as important as the technological advances is the powerful sense of awe and wonder still engendered by seeing films and pictures of men actually walking on and orbiting the surface of the Moon.

I therefore believe it is crucial for the advancement of our knowledge of Mars, the Solar System and our own planet that we plan to make direct human involvement an important part of space exploration, complementing the use of robot probes. Not to do so would be misguided and shortsighted in the extreme, and the NASA administrator's recent challenge to his engineers is to be welcomed, as would be a strong commitment to a return to the Moon. Let's hope the present US president has the foresight to support this bold and inspirational voyage.

Richard Hopkin

Department of Biological Sciences,
University of Durham,
Durham DH1 3LE, UK
e-mail: r.s.hopkin@dur.ac.uk

Time and motion

Sir — In her review of my book *Telling Time: Clocks, Diaries, and the English Diurnal Form, 1660–1785* (*Nature* **387**, 468; 1997), Kristen Lippincott warns readers that my thesis “seems to be somewhat flawed”. Her hesitancy is justified.

The flaw, she says, is this: I start from the observation that soon after Christiaan Huygens steadied clockwork's motion by inventing the pendulum regulator, the English language appears to have coined a phrase to describe the new clocks' sound: *tick, tick, tick* (not *tick-tock*, the onomatopoeia we have grown used to since its first appearance about 150 years ago).

Lippincott's objection is that clocks don't really sound like that. The escapement's alternate beats differ from each other more markedly than *tick, tick, tick* suggests.

That is true. I say so in the book: the “sameness” voiced by *tick, tick, tick* is

“fictional”, because “no successive impulses of the clock's escapement will actually sound identical” (page 8).

The phrase is the more striking because it is in part illusory; it emphasizes the new evenness of the intervals between the sounds over the quality of the sounds themselves. The first users of the phrase were (as Lippincott claims I am) “exaggerating to make a point”: that this was the way they newly imagined time to move — in small, steady increments — and that the new clocks were prompting them to imagine time this way.

Such imaginings have much to do, I try to show, with subsequent shapes of narrative, with the incremental, open-ended ways in which writers told stories of themselves and others, in the diaries, daily newspapers and other periodic forms that developed during the decades after Huygens' innovation.

Perceptions are not the same thing as reality, but they have history and consequences of their own, worth chronicling.

Stuart Sherman

Department of English,
Washington University,
Campus Box 1122,
One Brookings Drive,
St Louis,
Missouri 63130-4899, USA

Identity crisis

Sir — Joel E. Cohen's review of John Cairns' book *Matters of Life and Death* (*Nature* **387**, 565–566; 1997) refers to “seventeenth-century Breslau in Poland”.

In the seventeenth century, Breslau was the capital of the duchy of Silesia, which was part of the kingdom of Bohemia (and thus part of the Austrian empire). During the first Silesian war (also known as the War of the Austrian Succession) between Prussia (under King Friedrich II) and Austria (under the Empress Maria-Theresa), the duchy of Silesia became part of Prussia and remained so until the end of the Second World War, when the German population of Silesia was either killed or forced to leave and was replaced by Poles who had also been forced to leave their homes in eastern Poland.

J.-H. Klemme

Institut für Mikrobiologie & Biotechnologie,
Rheinische Friedrich-Wilhelms-Universität,
Meckenheimer Allee 168,
D-53115 Bonn, Germany

●The error was introduced in the *Nature* office. — Editor, *Nature*.

Risk of sea-change in the Atlantic

Stefan Rahmstorf

Emissions of greenhouse gases could weaken or even halt ocean overturning in the North Atlantic, radically altering the regional climate. It seems that the rate of greenhouse-gas increase may be as important as the final concentrations reached.

Will man-made climate change disrupt the ocean currents that have guaranteed Europe's mild climate for the past 10,000 years? A number of recent computer simulations as well as simple physical reasoning have hinted at such a possibility, but until now no systematic sensitivity study had been performed. On page 862 of this issue¹, Stocker and Schmittner present such a study, albeit with a highly simplified climate model. They conclude that ocean circulation stability depends not only on the total amount of greenhouse gas emitted by human activities, but also on the rate at which the gases are pumped into the atmosphere. It could make all the difference whether greenhouse gas emissions are reduced or continue unabated in the coming decades.

The reasons why climate researchers regard ocean circulation with concern are simple. The oceans transport massive amounts of heat around the planet. The northern North Atlantic in particular benefits from this — it receives around 10^{15} W of heat from the Gulf Stream and the North Atlantic Current (Fig. 1). This heat is released to the atmosphere and warms the winds that blow across Europe. The bulk of the heat is transported not by wind-driven ocean currents, but by the so-called thermohaline circulation, which is driven by temperature and salinity (and therefore density) differences in sea water. This circulation is basically a gigantic overturning motion, sometimes dubbed the 'ocean conveyor belt'. Warm surface waters flow north throughout the Atlantic, give off their heat and sink at high latitudes, and return south as cold water at a depth of about 2 km.

The crux of the matter is that the strength of the circulation, and thus the rate of heat transport, depends on small density differences, which in turn depend on a subtle balance in the North Atlantic between cooling at high latitudes and the input of less-dense fresh water from rain, snowfall and river runoff. More freshwater input would slow down the overturning, but not in a simple linear manner. Little happens at first, as the circulation continually removes the freshwater and replaces it with more salty water from the south. But there is a well-defined critical threshold — a saddle-node bifurcation, in

mathematical terms — beyond which the thermohaline circulation cannot cope with additional fresh water, and breaks down. (Surface warming can have a similar effect, although at high latitudes the sea water density is less sensitive to temperature.) The existence of this threshold was first proposed by Henry Stommel² in 1961 and has since been confirmed by a wide variety of ocean models, including global general circulation models³.

Deep ocean sediments and the Greenland ice cap contain a rich climate record which strongly suggests that the thermohaline circulation has broken down or at least changed drastically in the past after pulses of freshwater entered the Atlantic, and that this caused cold spells lasting for hundreds of years. The last of these events was the so-called Younger Dryas event 11,000 years ago. They do not necessarily give us indications for future events, as they occurred under ice age conditions, but they show that the possibility of a circulation breakdown is real. Global warming is expected to warm the surface waters and increase precipitation in high northern latitudes, both of which will reduce water density and move the Atlantic closer to the threshold. The crucial question is: how close?

Global warming simulations performed with coupled ocean–atmosphere circulation

models generally show that the Atlantic thermohaline circulation weakens by 15–50 per cent for a doubling of atmospheric carbon dioxide; Manabe and Stouffer⁴ found that after a quadrupling of carbon dioxide, the deep circulation in their model ground to a complete halt.

Such models would require more supercomputer time than is practical to explore the sensitivity of the ocean circulation to a wide range of parameters. That is why Stocker and Schmittner have used a highly simplified (but well-tested) model: a three-basin ocean model, averaged at each latitude, with simple atmospheric feedbacks. Although it is often said that 'the more feedbacks included, the more stable the model', the net effect of neglected feedbacks can be negative or positive; comparisons have shown that the ocean circulation in simple climate models can in fact be more stable than in the most sophisticated ones. Stocker and Schmittner's model gains credibility from the fact that it is consistent with the coupled circulation model of Manabe and Stouffer.

Nevertheless, a simple model like this cannot be expected to make accurate quantitative predictions. The key result of their study lies not in exact numbers, it is in the principle that the *rate* at which greenhouse-gas concentrations increase is crucial for the stability of the ocean circulation. Higher concentrations of greenhouse gases can be tolerated if they are approached more slowly. The implications for policy are clear: by starting to reduce emissions soon, we can buy greater climatic resilience and security later on.

Even the most sophisticated coupled climate models suffer from problems relevant to ocean circulation stability: *ad hoc* adjustments to the flux of heat and fresh water are often made in order to stop the models drifting to unrealistic situations, and these may artificially stabilize the ocean circulation; ocean convection is too small-scale to be

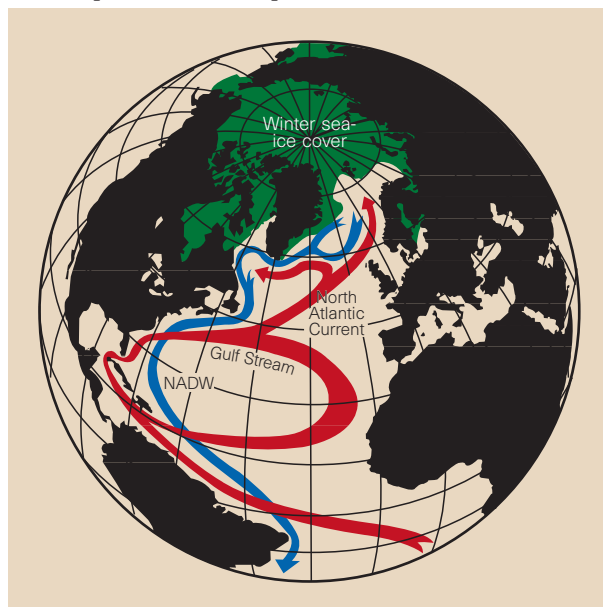


Figure 1 Europe's heating system. This highly simplified cartoon of Atlantic currents shows warmer surface currents in red and cold North Atlantic Deep Water (NADW) in blue. The thermohaline circulation heats the North Atlantic and Northern Europe. It extends right up to the Greenland and Norwegian Seas, pushing back the winter sea-ice margin. A rapid rise in greenhouse gas concentrations could disrupt the thermohaline circulation.

resolved explicitly in the models; and the physics of downslope flow of dense water near the sea bottom is not properly included. This last hampers the overflow of deep water over the sills between Greenland, Iceland and Scotland, so that too much of the models' deep water is formed south of the sills. Another source of uncertainty comes from the atmosphere models: changing precipitation is the major factor in changing ocean circulation, but it is much harder to predict than temperature changes.

The consequences of a disruption of the Atlantic thermohaline circulation are still under debate. Studies^{5,6} of the effects of an artificially triggered shutdown of the circulation, without accompanying global warming, typically find local sea surface cooling of 5–8 °C, with an even larger cooling in the atmosphere because increased sea-ice cover reflects more sunlight. Air temperature reduction is largest near Iceland, but affects much of Europe. In contrast, studies that show weakened or stopped circulation but include global warming find only a region of moderate cooling or reduced warming of the atmosphere, south of Greenland^{4,7,8}. Global warming can roughly compensate for the reduced oceanic heat transport in these experiments, because the ocean circulation winds down only slowly.

There are several caveats here. That the maximum effect is south of Greenland points at the overflow problem mentioned above. One can only speculate whether the response would be larger if the models formed more deep water north of the sills, so that the 'conveyor belt' would reach further north and interact with sea ice as it does in the real world. And a much faster circulation change (such as those seen in the ice-age climate records) may be possible through a different, convective type of instability³,

which has its own critical thresholds and depends on regional detail poorly represented in present climate models.

But whatever the effects on air temperature, such a change in ocean circulation would certainly have a severe effect on marine ecosystems and fisheries. Even small fluctuations in ocean currents have led to the collapse of fish stocks and sea-bird populations in the past. Another concern is that a reduced thermohaline circulation would weaken the carbon dioxide uptake of the ocean⁹, effectively making the climate system more susceptible to anthropogenic emissions.

So a collapse of the Atlantic thermohaline circulation would probably have serious consequences, involving risks that no nation bordering the North Atlantic would willingly take. Climate models are still too coarse to accurately predict how vulnerable the ocean circulation is, but they suggest that crossing a critical limit is within the range of possibilities for the next century. A disruption of the thermohaline circulation cannot be ruled out if we continue to pollute the atmosphere at the present rate. The work of Stocker and Schmittner is a timely reminder, before the Kyoto climate summit in December, that swift action is needed to reduce the risk of unwelcome climatic surprises. □

Stefan Rahmstorf is at the Potsdam Institute for Climate Impact Research (PIK), PO Box 60 12 03, D-14412 Potsdam, Germany.
e-mail: rahmstorf@pik-potsdam.de

1. Stocker, T. & Schmittner, A. *Nature* **388**, 862–865 (1997).
2. Stommel, H. *Tellus* **13**, 224–230 (1961).
3. Rahmstorf, S. *Nature* **378**, 145–149 (1995).
4. Manabe, S. & Stouffer, R. J. *Nature* **364**, 215–218 (1993).
5. Manabe, S. & Stouffer, R. J. *J. Clim.* **1**, 841–866 (1988).
6. Schiller, A., Mikolajewicz, U. & Voss, R. *Clim. Dyn.* **13**, 325–347 (1997).
7. Cubasch, U. et al. *Clim. Dyn.* **11**, 71–84 (1995).
8. Hirst, A. C., Gordon, H. B. & O'Farrell, S. P. *Geophys. Res. Lett.* **23**, 3361–3364 (1996).
9. Sarmiento, J. L. & Le Quéré, C. *Science* **274**, 1346–1350 (1996).

Biological hydrostats

Wrapping the armadillo's penis

Richard Wassersug

A multitude of soft-bodied organisms (such as worms) and parts of organisms (such as the tube feet of echinoderms, the trunks of elephants, and even our own tongues) are biological hydrostats. They lack solid skeletons, yet hold their shape because of the internal fluid pressure, which is resisted by tension in the sheath that surrounds the fluid space. These enwrapping sheaths are reinforced with fibres that classically spiral up and down the long axis of the structure, forming a crossed helical array.

The penis is also a biological hydrostat, and in the *Journal of Morphology* Diane A. Kelly¹ reveals that, in the nine-banded armadillo (*Dasypus novemcinctus*; Fig. 1), it

is reinforced by investing fibres that run, not helically, but orthogonally to the long axis of the penis. That is, the fibres run lengthwise and circumferentially (Fig. 2, overleaf), and the penis is the first biological structure known to have such a pattern.

The mechanical properties of biological hydrostats are greatly affected by the angle of the supporting fibres within the sheath². When the fibres cross the long axis of the structure at about 55° — which they typically do — the structure can easily bend without buckling, while it retains a constant volume. This is fine for worms and tongues, but what about hydrostats that need the opposite physical properties? Suppose one has a biological structure that must be able to increase

its volume enormously, and then resist any bending or long-axis compression. Such a structure is the mammalian penis.

The supporting fibres in the armadillo penis are made of collagen and lie in the tunica albuginea, the thin layer of fibrous tissue that surrounds the erectile tissue (corpus cavernosum). Kelly¹ shows that fibres in the armadillo's tunica are precisely orientated at either 0° or 90° to the long axis of the penis. When the penis is flaccid the fibres are massively pleated. These pleats allow for the great increase in the length and width of the penis during erection, and they disappear when the penis is fully tumescent.

The mechanical properties of a hydrostat with an axial orthogonal array (such as that of the penis) were understood by biologists even before they realized that there were biological structures with this pattern². An orthogonal array, unlike a helical one, provides the penis with maximum flexural stiffness when erect. This is essential for effective intromission — the armadillo whose penis buckles when it should not, is left out in the cold without descendants.

If bending loads on the erect penis are excessive, the longitudinal fibres can rupture, leading to a fractured penis. This can happen to humans if the penis is exposed to blunt trauma³, although it is thankfully rare because of the enormous strength of the collagen array in the tunica albuginea. Whether armadillos in nature ever fracture their penises is not known. But Kelly was able to reproduce the situation by inflating armadillo penises in the laboratory and testing their strength under various loading regimes. She confirmed that they failed by buckling, under the same loading regimes as those that lead to fracture of the penis in humans.

From Kelly's unpublished survey of other mammalian species, it is likely that all mammalian penises are wrapped the same way as the armadillo's, in orthogonal-fibre arrays. So why has it taken us so long to realize this? One can only surmise that most scientists (or the male ones at least) have been



Figure 1 Well hung — the nine-banded armadillo, *Dasypus novemcinctus*.

JEFF FOOT/BBC NATURAL HISTORY

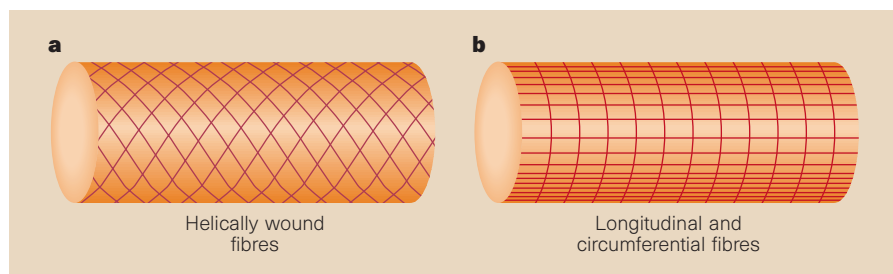


Figure 2 Biological hydrostats hold their shape because the pressure of the fluid inside is resisted by tension in the surrounding sheath. a, Most biological hydrostats, such as earthworms, are reinforced by helically wound supporting fibres. b, Diane Kelly¹ has now shown that the penis of the nine-banded armadillo is reinforced by longitudinal and circumferential fibres, and this orthogonal orientation is likely to be the same for all mammals. (Adapted from ref. 2.)

reluctant to take on such a project for either social or psychological reasons.

Because the orthogonal orientation of the collagen fibres in the tunica albuginea is probably the same for all mammals, one can't help but wonder why Diane Kelly chose to study the fibres in such an off-beat animal as the armadillo. The answer seems to be simple convenience — she collected her specimens as fresh road-kills along the highways of northern Florida. Armadillo carcasses are abundant there because of the animal's 'defence' against a perceived threat. It launches itself vertically into the air when frightened, which all too often means that the armadillo flies directly into the grills of cars that might otherwise pass over it.

It is also true that armadillos have exceptionally large penises. According to Kelly, the penis of the nine-banded armadillo, when fully erect, is about one-third the

length of its body. However, in other armadillo species it can be two-thirds the length of the animal.

The practical implications of Kelly's discovery have yet to be explored. Could it lead to lighter, stronger yet safer condoms? Could it improve the design of penile implants? To apply these results to clinical situations relies on a knowledge of the loads to which the penis is subjected during normal function. Collecting such data will be a challenge for Kelly — or anyone else who pursues this field of research. □

Richard Wassersug is in the Department of Anatomy and Neurobiology, Dalhousie University, Sir Charles Tupper Medical Building, Halifax, Nova Scotia, Canada B3H 4H7.

1. Kelly, D. A. *J. Morphol.* **233**, 249–255 (1997).

2. Vogel, S. *Life's Devices: The Physical World of Animals and Plants* (Princeton Univ. Press, Princeton, NJ, 1988).

3. Koga, S. *et al. Br. J. Urol.* **72**, 228–229 (1993).

Quantum mechanics

Atoms that agree to differ

D. Bouwmeester and A. Zeilinger

In our everyday world, things have properties independent of whether we care to look at them or not. Whether a given apple is red or green is independent of our checking its colour. And although most people acknowledge that quantum mechanics is very strange, they often feel that quantum objects still have their properties — it seems to be just the clumsiness of our tools that invariably disturbs quantum objects in such a way that we cannot observe all their properties. But any seasoned quantum mechanic knows this not to be true. For example, physically separated objects can become 'entangled' in a collective state, each possessing no well-defined state of its own. Experimental observation of entanglement so far has been restricted to photons¹, electrons and protons. But Hagley *et al.*² have now demonstrated entangled pairs of atoms. In their experiment they produce pairs of atoms, one in the ground state and the other in an excited state, but it is completely undefined which is which.

Hagley and colleagues use a beam of rubidium atoms prepared in so-called circular Rydberg states. These are highly excited states in which the outermost electron is confined to an effectively circular orbit, very much like planets around the Sun. They use states with principal quantum number 50 and 51; the first they call the ground state for the experiment and the second one the excited state.

The authors first prepare an atom in an excited state, then send it through a microwave cavity which resonates at the same energy as the transition from the excited to the ground state. This means that the atom can gradually make the transition by exciting the microwave field inside the cavity, just as the correct sound frequency can bring a string to resonance. If the atom stays inside the cavity long enough, there comes a point at which all its energy has been transferred to the cavity, and we can say that the atom has emitted a photon into the cavity. But if the

atom spends less time inside the cavity, we cannot know whether or not it has emitted a photon — we can only assign a probability to that event.

In the new experiment, the time that the atom spends inside the cavity is chosen so that if we were to make a measurement of the state of the cavity field afterwards, we would have 50% chance of observing the photon there. Observing the photon would imply that the atom is now in its ground state; observing no photon would imply that the atom is still in its excited state. But in the experiment, no measurement is performed on the cavity field, and that leaves the cavity-atom system in a superposition of these two possibilities. This superposition of states in a composite quantum system is called an entangled state.

In order to entangle two atoms, a second atom, prepared in its ground state, passes through the cavity. (The cavity is made of superconducting niobium, operating at a very low temperature in order to isolate the cavity field well enough from its surroundings to preserve its superposition of states.) This second atom spends enough time inside the cavity for the transition from ground state to excited state to happen if a photon were inside the cavity. If no photon were present the second atom would remain in its ground state. In fact, there is a superposition of photon and no photon so, once the second atom leaves the cavity, the entanglement between the cavity and the first atom is transferred to entanglement between the two atoms.

The entanglement implies that, after the interaction with the cavity, all that the two atoms know is that they have to be different — one has to be excited the other one has to be in the ground state — but neither knows what state it should be in. Subsequently measuring one atom defines its state, and so forces the other atom into a well-defined (opposite) state.

Because the two atoms are separated in space, the measurement process has non-local features — that is, some 'influence' could be said to be travelling faster than light (in fact, instantaneously) between the two atoms. This might seem to be in conflict with the theory of relativity, and such arguments were at the root of criticism by Einstein and many others of the foundations of quantum mechanics. (Even though the non-local features of entangled states can by no means be exploited to send information faster than the speed of light.) In fact, the experiment performed by Hagley *et al.* is in the spirit of a thought experiment considered by Einstein, Podolsky and Rosen to demonstrate the absurdity of quantum mechanics. It confirms that the rules of quantum mechanics are valid, no matter how absurd they might seem.

The rebellious nature of entangled systems has been made a virtue in the continu-

ing development of quantum computation, quantum cryptography and quantum teleportation. It is entanglement that makes a quantum computer, in principle, so much more powerful than a classical computer; that makes two-photon cryptography 100% safe; and that makes teleportation of quantum states possible. Quantum computation is far from being experimentally realized, but the production of two entangled atoms is a small but important step towards this goal.

In a parallel development, Ed Fry's group in Texas is trying to create entangled states of atoms by dissociating molecules using weak lasers³. And one might speculate about larger objects: there is no reason not to expect entanglement to be observed between molecules, and it is very hard to say what the ultimate size limit will be. The experimental

problem is that the larger an object is, the easier it can be disturbed by minute external influences — the quantum state describing many entangled particles is extremely fragile. But there is no reason in principle not to expect entanglement of macroscopic objects. As we know from the discussion between Albert Einstein and Niels Bohr in the early days of quantum mechanics, the distinction between quantum and classical is not necessarily the distinction between microscopic and macroscopic. □

D. Bouwmeester and A. Zeilinger are at the Institut für Experimentalphysik, Universität Innsbruck, Technikerstrasse 25, A-6020 Innsbruck, Austria.

1. Aspect, A., Grangier, P. & Roger, G. *Phys. Rev. Lett.* **47**, 460–463 (1981).
2. Hagley, E. *et al.* *Phys. Rev. Lett.* **79**, 1–5 (1997).
3. Fry, E. S., Walther, T. & Li, S. *Phys. Rev. A* **52**, 4381–4395 (1995).

Genomic stability

Silencing and DNA repair connect

Stephen P. Jackson

One of the challenges facing molecular biologists is to explain how the huge amount of DNA within the eukaryotic cell nucleus is packaged into chromatin, of which chromosomes are composed, and how this packaging affects processes such as DNA replication, transcription and the maintenance of genomic integrity. On page 900 of this issue¹, Tsukamoto, Kato and Ikeda reveal that three yeast proteins, associated with chromatin and with transcriptional silencing, are also involved in the repair of DNA damage induced by ionizing radiation. These results provide an unexpected dimension to silencing and chromatin organization, and may yield clues to the mechanism

by which DNA double-strand breaks are rejoined.

The fundamental unit of chromatin is the nucleosome, in which DNA is wound around an octamer of the core histones H2A, H2B, H3 and H4. Interactions with linker histones and various non-histone proteins provide further DNA compaction, which, in defined regions of the genome, culminates in the highly condensed state termed heterochromatin. Notable features of heterochromatin are that it is generally transcriptionally silent, is replicated late in S-phase of the cell cycle, is localized to the nuclear periphery, and is relatively inaccessible to DNA-modifying enzymes².

Heterochromatin-mediated transcriptional silencing is most commonly observed in higher eukaryotes, but it also operates in less complex organisms. For example, in the yeast *Saccharomyces cerevisiae*, the factors Sir2, Sir3 and Sir4 function in transcriptional silencing at telomeres and ensure that genetic information in the silent mating cassettes is not inappropriately expressed. All three factors are believed to be targeted to these loci by proteins such as Sir1 and Rap1 (ref. 2). Although the exact mechanism of silencing is unknown, the ability of Sir3 and Sir4 to interact with the amino-terminal tails of histones H3 and H4 (ref. 3), and the requirement of these tails for effective transcriptional silencing, suggest that the Sir proteins help to establish a higher-order, repressed state of chromatin.

The linkage between Sir proteins and DNA repair now reported by Tsukamoto *et al.*¹ came to light through investigations into Ku, a protein that binds to DNA double-strand breaks. This factor functions in the repair of radiation-induced DNA strand breaks by the DNA non-homologous end-joining pathway^{4,5}. It is also required for V(D)J recombination, which helps to generate the virtually limitless antigen-binding repertoire of the vertebrate immune system. The prevailing view is that Ku binds to DNA strand-breaks and recruits other factors involved in DNA repair and DNA damage signalling.

Using the yeast two-hybrid screen, Tsukamoto *et al.* discovered that the *M*, ~70,000 subunit of yeast Ku, named Hdf1 or Yku70, interacts with the carboxy-terminal portion of Sir4. Furthermore, they reveal that inactivation of the *SIR2*, *SIR3* or *SIR4* genes (but not *SIR1*) leads to severe defects in Ku-dependent DNA end-joining, but does not affect the homologous recombination pathway for DNA double-strand break repair. Taken together with radiation sensitivity data, these results show that Sir2, Sir3 and Sir4 are components of the Ku-associated DNA repair apparatus. Given the strong evolutionary conservation of Ku, it is tempting to speculate that functional homologues of the yeast Sir proteins exist in mammalian systems, and that they take part in V(D)J recombination and the repair of DNA damage induced by ionizing radiation.

As well as interacting with yeast Ku70 in the two-hybrid screen, Sir4 also binds to several other factors, including histones H3 and H4, and Rap1 and Sir3 (refs 3, 6). These observations imply that certain interactions with Sir4 could be mutually exclusive, and could mean that the Sir4–Ku70 interaction is indirect. Nevertheless, from the current work it seems reasonable to suppose that Ku recruits the Sir protein complex to sites of DNA damage, and that this might lead to the establishment of a heterochromatin-like state. ►

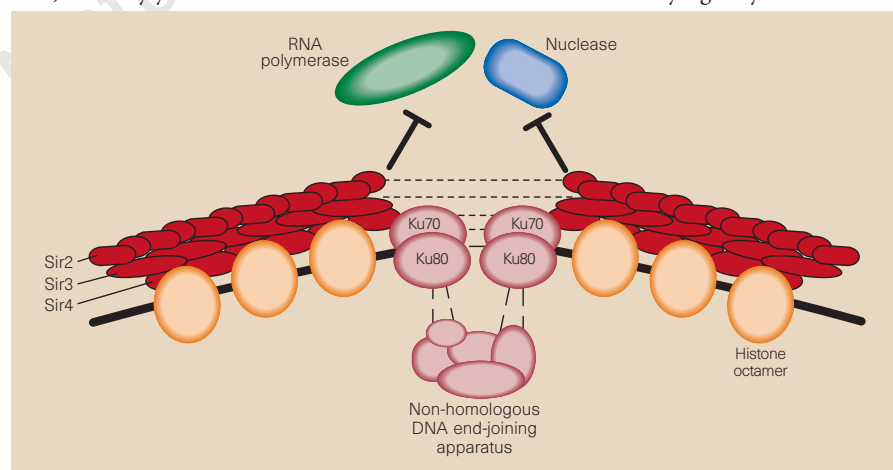


Figure 1 The work by Tsukamoto *et al.*¹ shows that, together with the DNA end-binding factor Ku, yeast Sir2, Sir3 and Sir4 proteins operate in rejoining breaks in double-stranded DNA. Shown here is a hypothetical mechanism for Ku-mediated assembly of Sir2, Sir3 and Sir4 (red) at the site of a DNA break. Both Ku subunits are shown; dashed lines represent protein–protein interactions that might tether the two DNA ends together and recruit the non-homologous DNA end-joining apparatus. The DNA is depicted as a thick black line, and the histone octamer as a single orange oval for simplicity. As indicated, this model for DNA repair involves the prevention of transcription and of DNA degradation by nucleases.

There are several ways in which such a process could potentiate DNA repair (Fig. 1). First, because silenced chromatin is generally inaccessible to DNA-modifying enzymes, it may prevent nuclease-mediated degradation of damaged DNA ends. Second, heterochromatin-associated silencing could stop processes such as transcription from interfering with DNA repair. Third, heterochromatin-mediated condensation of the damaged DNA might facilitate the juxtaposition and subsequent ligation of the two broken DNA molecules, and could prevent these DNA ends from engaging in other recombination reactions. This third function could be particularly important *in vivo*, where damaged DNA ends might otherwise easily become irrevocably separated, and where recombination with other loci is generally undesirable.

The new data also raise the possibility that the Sir proteins target Ku to certain heterochromatic regions in the absence of DNA damage. Consistent with this idea, as is the case for inactivation of *SIR3* or *SIR4* (ref. 7), loss of yeast Ku function leads to telomeric shortening^{8,9}. Thus, it will be of interest to determine whether Ku affects other Sir-dependent functions, such as tethering

telomeres to the nuclear periphery and transcriptional silencing (in this regard, it is noteworthy that mammalian Ku has been shown to suppress transcription⁵). Finally, the relative inaccessibility of heterochromatin may mean that it is inherently difficult to repair DNA damage in these regions of the genome — in such chromosomal contexts, Sir-mediated targeting of Ku could be a mechanism for potentiating the DNA non-homologous end-joining system to resolve potentially lethal DNA damage. □

Stephen P. Jackson is at the Wellcome/CRC Institute, and the Department of Zoology, University of Cambridge, Tennis Court Road, Cambridge CB2 1QR, UK.

e-mail: spj13@mole.bio.cam.ac.uk

1. Tsukamoto, Y., Kato, J.-i. & Ikeda, H. *Nature* **388**, 900–903 (1997).
2. Grunstein, M. *Curr. Opin. Cell Biol.* **9**, 383–387 (1997).
3. Hecht, A., Laroche, T., Strahl-Bolsinger, S., Gasser, S. M. & Grunstein, M. *Cell* **80**, 583–592 (1995).
4. Lieber, M. R., Grawunder, U., Wu, X. & Yaneva, M. *Curr. Opin. Genet. Dev.* **7**, 99–104 (1997).
5. Jackson, S. P. & Jeggo, P. A. *Trends Biochem. Sci.* **20**, 412–415 (1995).
6. Moretti, P., Freeman, K., Coodly, L. & Shore, D. *Genes Dev.* **8**, 2257–2269 (1994).
7. Palladino, F. *et al.* *Cell* **75**, 543–555 (1993).
8. Porter, S. E. *et al.* *Nucleic Acids Res.* **24**, 582–585 (1996).
9. Boulton, S. J. & Jackson, S. P. *Nucleic Acids Res.* **24**, 4639–4648 (1996).

RNA structure

A molecular contortionist

Jennifer A. Doudna

RNA molecules kink, bend, loop and twist themselves into a wonderful variety of shapes. These took centre stage in June, at the first meeting devoted solely to RNA structure*. Due to the abundance of biologically important RNAs and the development of technical advances to study them, RNA structural biology has now come of age. And it is clear that ribonucleic acids are capable of a level of structural complexity that was once thought to be confined to proteins.

Although RNA lacks the chemical diversity of proteins, the many conformational degrees of freedom of its phosphate backbone, along with the unique hydrogen-bonding and stacking properties of its nucleotide bases, provide ample resources for three-dimensional folding. This has been exploited by evolution — RNA molecules are integral parts of the cellular machinery for protein biosynthesis, RNA processing, chromosome end replication and protein transport (Table 1). RNA molecules are sometimes catalysts (ribozymes), and the functional diversity of this molecular jack-of-all-trades implies structural complexity.

At first glance, the principles of RNA folding seem to be simple and rather dull,

because the four nucleotide building blocks readily form hydrogen-bonded base pairs and a corresponding double helix. But, unlike its DNA sibling, RNA rarely exists as a long, straight duplex. Instead, many RNA molecules fold back on themselves to produce short stretches of base-pairing interspersed with unpaired 'internal loops', bulges and terminal loops. And these loops and bulges are the key to the formation of complex RNA structures.

Internal loops, it turns out, are surprisingly ordered — the loop nucleotides make unconventional, yet specific, interactions.

For example, according to the nuclear magnetic resonance (NMR) and crystal structures of the loop-E region of 5S ribosomal RNA, purine nucleotides within the 'loop' are pinched together in sheared base-pairs, the bases of one strand stacking on bases in the opposite strand (P. Moore and T. Steitz, Yale Univ.). Although the overall structure is helical, the major groove of the loop region is stapled shut by magnesium ions that bridge the phosphate backbones.

It is not known whether this dramatically distorted geometry is recognized by the ligand for the loop-E region, the L25 ribosomal protein. But other RNA internal loops change shape on binding their ligands, for example the Tat and Rev responsive-element RNAs of the human immunodeficiency virus (HIV) and an RNA that binds the small molecule theophylline (A. Pardi, Univ. Colorado). The NMR structure of the unbound tetraloop receptor — an internal loop that binds to GAAA tetraloops and is common in large RNAs — reveals a 'closed' helical conformation, containing cross-strand purine stacking that resembles the loop-E structure (J. Feigon, UCLA). When the tetraloop is docked, however, the receptor helix is opened to allow stacking of the tetraloop bases in the minor groove.

Localized regions of RNA can be dynamic, perhaps providing the basis for movement on a grand scale within large RNAs. Crosslinking of active conformers of the *Tetrahymena* self-splicing intron, for example, reveals that helices within this roughly 400-nucleotide RNA can move by as much as 60 Å (T. Cech, Univ. Colorado). In the wish-bone-shaped hammerhead ribozyme (Fig. 1), movement of the helical arms is probably more limited, but it may be required to position an active-site magnesium ion close to the labile bond at the moment of cleavage (D. Herschlag, Stanford Univ.).

Metal ions are central to RNA structure and catalysis — a point that emerged as a major theme of the meeting. For example, more than 100 magnesium ions are needed for the several hundred nucleotides of ribonuclease P to be folded (C. Fierke, Duke

Table 1 The ABCs of RNA taxonomy

RNA ABCs	Description
cRNA	Ribozymes and autocatalytic molecules
gRNA	Template, or guide sequences, for editing of RNA messages
mRNA	Messenger that carries information from DNA to be translated into protein
rRNA	Essential components of the ribosome, the protein biosynthetic machinery
snRNA	Small nuclear components of the mRNA splicing machinery called the spliceosome
snoRNA	Small nucleolar RNAs that specify methylation sites in rRNA, and may have other chaperone-like functions
tRNA	Transfer molecules that carry amino acids
Other examples:	
Telomerase RNA	Chromosome end-replication template
Signal-recognition particle (SRP) RNA	Essential component of SRP, involved in protein transport in cells
Xist RNA	Required for X-chromosome inactivation during development

*RNA Structure, Univ. California, Santa Cruz, 25–29 June 1997.

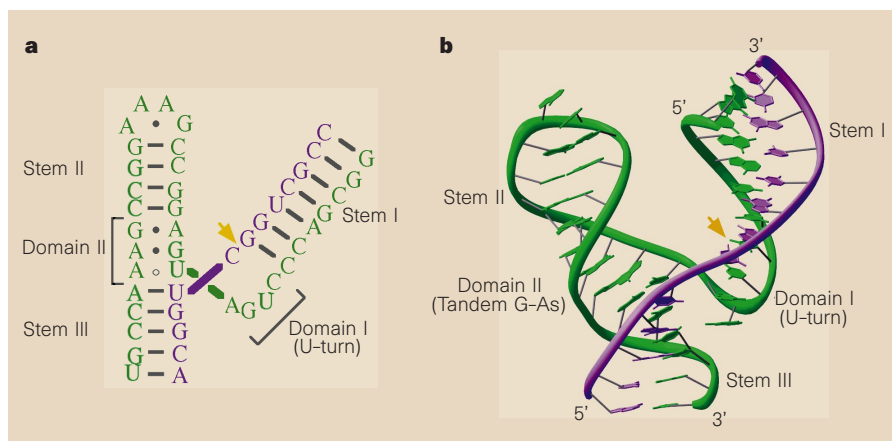


Figure 1 The hammerhead ribozyme, found in plant-virus-like RNAs, self-cleaves at the phosphodiester bond indicated by the arrow. a, Secondary structure; b, three-dimensional shape, as revealed by X-ray crystallography.

Univ.). And in the P4–P6 domain of the *Tetrahymena* self-splicing intron, a cluster of magnesium ions binds to phosphate oxygens and guanosine bases to form a structural core. Magnesium ions may play a similar role in the folding of a three-helix junction in ribosomal RNA (J. Williamson, MIT), and they can also bind the major groove of RNA helices near G–U base pairs. Group-II self-splicing RNAs may use this trick to position active-site metals (A. M. Pyle, Columbia Univ.).

Unique structures in RNA provide handles for proteins, yet the specificity of these interactions has been a mystery. How are transfer RNAs recognized by elongation factors and cognate tRNA synthetases, for example, at the appropriate stages during protein biosynthesis? X-ray crystallographic studies reveal that these proteins ‘see’ different features of tRNAs: elongation factor Tu (EF-Tu) sees the aminoacylated tRNA acceptor stem (P. Nissen, Aarhus Univ.); the histidyl tRNA synthetase recognizes a motif within the acceptor stem of tRNA-His (R. Giegé, CNRS, Strasbourg); and seryl-tRNA synthetase sees a variable hairpin loop of tRNA-Ser (S. Cusack, EMBL, Grenoble).

In the case of the spliceosomal proteins U1A and U2B'', specificity for RNA loops that differ by just one nucleotide is conferred by the interaction of U2B'' with a second protein, U2A' (K. Nagai, MRC). Single-stranded-RNA binding proteins, including U1A, U2B'' and the MS2-phage coat protein, use a β -sheet motif to recognize splayed nucleotides in the RNA (F. Allain, UCLA; K. Nagai; L. Liljas, Uppsala Univ.). But a double-stranded-RNA binding domain from *Xenopus* grabs its substrate by interacting with two successive minor grooves along one face of the RNA duplex (S. Schultz, Univ. Colorado). Localized distortion widens the major groove of the target RNA, and a network of ordered water molecules lines the protein–RNA interface.

The emerging themes of RNA structure—stable loops, major- and minor-groove inter-

actions, modular motifs and metals—indicate that it may one day be easier to model and predict RNA folds than to predict protein structure. New and modified algorithms for identifying secondary structures in RNA sequences are already used. Moreover, a database of structural models and corresponding biochemical data for the 30S ribosomal subunit is now available on the World Wide Web (<http://www-smi.stanford.edu/projects/helix/riboweb.html>) (R. Altman, Stanford Univ.).

Modelling of RNA tertiary structures has proven more difficult. Although a model of the HIV Rev-binding element compared favourably to an NMR structure (F. Leclerc, Univ. Montreal), a model of the ribosomal A-site (S. Harvey, Univ. Alabama) was quite different from the NMR structure of this region (J. Puglisi, Univ. California, Santa Cruz). Models of the hammerhead and other larger ribozymes continue to spark new ideas about catalytic mechanisms (E. Westhof, CNRS).

One of the most exciting directions of RNA structural biology is now to understand how large RNAs, and RNA–protein complexes, work. Chemical cross-linking and phylogenetic analyses have been used to model group-II self-splicing introns (F. Michel, CNRS), ribonuclease P RNA (N. Pace, Univ. California, Berkeley) and the ribosome (H. Noller, Univ. California, Santa Cruz). But a detailed understanding of these RNAs will require structures at atomic resolution.

Tantalizing images of ribosome crystals were presented during the final hour of the meeting (H. Noller and C. Wilson, Univ. California, Santa Cruz; T. Steitz). So when the RNA structural biologists reconvene in two years, honoured guests may include, among other RNA behemoths, the ribosome itself—an exciting prospect indeed. □

Jennifer A. Doudna is in the Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06520, USA. e-mail: doudna@csb.yale.edu



100 YEARS AGO

It has long been my conviction that we study animals too much as dead things. We name them, arrange them according to our notions of their likeness or unlikeness, and record their distribution. Then perhaps we are satisfied, forgetting that we could do as much with minerals or remarkable boulders. Of late years we have attempted something more; we now teach every student of Zoology to dissect animals and to attend to their development.... But the animals set before the young zoologist are all dead; it is much if they are not pickled as well. When he studies their development, he works chiefly or altogether upon continuous sections, embryos mounted in balsam, and wax models. He is rarely encouraged to observe live tadpoles or third-day chicks with beating hearts. As for what Gilbert White calls the *life and conversation of animals*, how they defend themselves, feed, and make love, this is commonly passed over as a matter of curious but not very important information; it is not reputed scientific, or at least not eminently scientific. – L. C. Miall

From *Nature* 26 August 1897.

50 YEARS AGO

In facing realities we must agree that the world is not yet at peace, neither is it free from the threat of further conflagration, perhaps even on a world-wide scale. It follows that though every effort towards world peace must be made, it would be foolish during present troublesome times to fall into a state of lethargy and unpreparedness. So some scientific research, especially that dealing with certain aspects of atomic energy, must even now, and perhaps for some time to come, be carried on under the ban of official secrecy. Feeling is high these days, and concern is often expressed that military secrecy might take the place of former industrial interests in slowing up the free flow of ideas which are essential to the smooth progress of scientific endeavour. But at present men of science must face the fact that some secrecy is inevitable, and this makes it all the more imperative, therefore, that they themselves should strive to bring before each other and above all before the general non-scientific public the many beneficent aspects of science.

From *Nature* 30 August 1947.

Autoimmune diabetes

Retrovirus as trigger, precipitator or marker?

Christophe Benoist and Diane Mathis

Insulin-dependent diabetes mellitus (IDDM) is an autoimmune disease that is characterized by the specific destruction of insulin-producing β -cells in the pancreatic islets of Langerhans^{1,2}. Despite decades of intensive investigation, the trigger for this self-attack has remained a mystery. But in the 25 July issue of *Cell*, Conrad *et al.*³ describe new findings which suggest that a virus may be responsible.

It has long been thought that human IDDM may be triggered by an infectious agent^{1,2}. The main evidence is that the concordance rate for monozygotic twins is only about 50 per cent; that incidence varies along a north-south gradient; and that the frequency of diabetes is increasing at a striking rate in some countries. But there are alternative explanations for these epidemiological phenomena — in fact, most of the variations within and between countries (for example, the higher incidence in more northern European countries) fit better with the idea that infections nonspecifically protect from,

rather than incite, disease. This idea would also be more consistent with observations on the non-obese diabetic mouse (the most popular animal model of autoimmune diabetes), which has the highest incidence of disease when it is housed under germ-free conditions. So, because there has never been direct evidence that an infectious agent triggers human IDDM, support for the hypothesis of an infectious trigger dwindled.

In 1994, Conrad *et al.*⁴ reported intriguing results that rekindled the debate on the involvement of a virus or bacterium in IDDM. They analysed the repertoire of T lymphocytes that had invaded the pancreatic islets of two recently deceased, acutely diabetic patients. According to information from animal models^{1,2}, T cells are the primary mediators of IDDM. The authors found that T cells whose antigen receptors (T-cell receptors; TCRs) carried the V β 7 variable region were strikingly enriched in these patients. This over-representation was peculiar to the infiltrating cells, because it

was not observed with T cells in the blood of the same patients.

Such an enrichment for a particular V β region evokes the behaviour of superantigens. These proteins, produced by certain viruses and bacteria, can stimulate an inordinately large number of T cells. They do this by binding directly to a subset of V β regions, rather than by interacting in the usual, much more restricted, fashion with a TCR α/β -chain combinatorial site. On the basis of these results, Conrad and colleagues proposed that viral or bacterial superantigens are the trigger for IDDM. But the significance of these findings was questioned on several counts, most notably because only two, rather atypical, patients were involved in the study. Moreover, the predicted superantigen — and whatever produced it — needed to be identified⁵.

In their new paper, Conrad *et al.*³ describe the isolation of a hitherto unknown endogenous retroviral genome from a patient with diabetes, and the identification of a fragment of one of its proteins as a superantigen. Given the lack of direct evidence for the action of an exogenous infectious agent as the trigger for human IDDM, the authors surmised that an endogenous retrovirus might be involved. Vertically transmitted retrovirus genomes (both complete and partial) are found in abundance in the genomes of vertebrates, and some of them encode potent superantigens. Moreover, the expression of particular endogenous retrovirus transcripts specifically in pancreatic-islet β -cells, has been correlated with disease in non-obese diabetic mice^{6,7}.

To assay for a culprit retrovirus, Conrad *et al.* measured virus-derived reverse-transcriptase activity in supernatants from cultures of freshly isolated pancreatic islets. They found activity in the samples from the two patients mentioned above, but not in those from non-diabetic people. The reverse transcriptase seemed to derive from leukocytes (such as B cells, T cells, macrophages and dendritic cells) rather than from β -cells, because supernatants from cultures of splenocytes had even more activity. The authors then used a sophisticated polymerase-chain-reaction strategy to isolate a complete retroviral genome from the islet supernatant of one of the diabetics. Transcripts derived from (or at least related to) this genome were found in blood from 10 diabetic patients, but not from non-diabetic controls. The retrovirus had never previously been identified, but it shared homology with mouse mammary tumour viruses, which produce superantigens. So they searched for — and found — a superantigen encoded by the newly identified viral genome, located within the gene for the envelope protein. Suggestively, this superantigen stimulated T cells that displayed TCRs with V β 7, but not those that presented other variable regions. ►

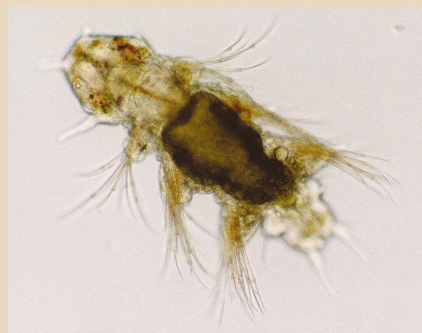
Cryobiology

Cold comfort for anglers and toxicologists

The ragworm *Nereis virens* is an object of some desire, for not only is it widely favoured as bait by anglers but it is also used in toxicity assays. Demand for *Nereis* outstrips supplies from natural resources, however, not least because breeding is highly seasonal.

Hence the interest in cryopreservation of ragworm larvae and in the specimen pictured here, which has been down to -196°C (the temperature of liquid nitrogen) and back again. The larva emerged from the process alive, well and ready for rearing to adulthood to do its commercial duty. The feat is reported by P. J. W. Olive and W. B. Wang of the University of Newcastle upon Tyne (*Cryobiology* 34, 284–294; 1997), who describe the painstaking permutation of experimental protocols and age of larvae necessary to achieve high rates of successful cold preservation and recovery of larvae. They are able, they say, to preserve over a million larvae from a single artificial fertilization of *Nereis*.

The technique is subject to a patent application by Seabait Ltd. It hinges on the rate of the initial cooling of the larvae to around -35°C before they are plunged into liquid nitrogen: too fast, and ice formation



will occur inside the larvae and kill them; too slow, and the cryoprotectant employed (in this case dimethyl sulphoxide) will poison them.

Cryopreservation methods are well established for gametes and tissues for transplantation, and for some years have been applied to the early embryonic stages of certain insects and mammals. But Olive and Wang believe that, apart from organisms that are naturally cold-resistant, the young *Nereis* they use are the most developmentally advanced multicellular creatures to have been successfully subjected to the process. The larvae have functional muscle, nervous and digestive systems, and eyes and other sense organs, and they can crawl or swim. **Tim Lincoln**

Conrad *et al.* propose a model to explain how this endogenous retrovirus provokes diabetes (Fig. 1a). For some reason, the retroviral genome starts to be transcribed in leukocytes, giving rise to superantigen. The synthesis of superantigen in 'professional' antigen-presenting cells that express major histocompatibility complex class II molecules provokes a broad, systemic response by most of the T lymphocytes that display particular V β s — especially V β 7. By chance, some of the activated T cells can also respond to antigens that are specifically found in pancreatic-islet β -cells. Before their activation by superantigen, these T cells do not attack the β -cells because, being naive, they cannot circulate through tissues.

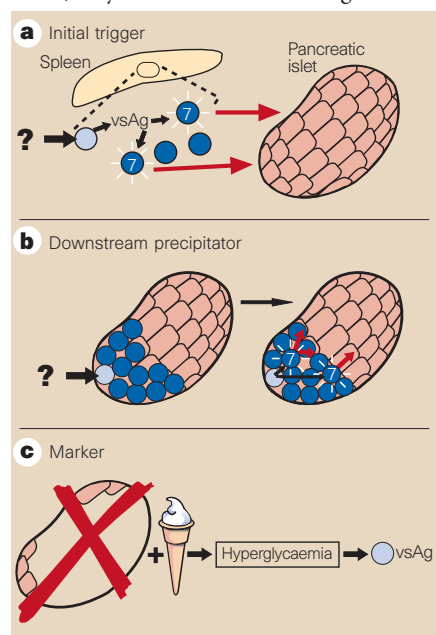


Figure 1 Conrad *et al.*³ have isolated the complete genome of a hitherto unknown endogenous retrovirus in the supernatant of pancreatic-islet cultures from a diabetic patient. a, The authors believe that the retrovirus is the initial trigger for diabetes. An unknown event (stress, infectious agent, hormonal perturbations) induces activation of the endogenous retrovirus in the leukocytes found in lymphoid organs. The retrovirus-encoded superantigen (vsAg) activates V β 7-positive T cells (7), which can then travel to non-lymphoid organs. By mischance, some of the V β 7-positive T cells are reactive to pancreatic antigens, which home to the pancreas and destroy the islets. b, An alternative explanation is that the same unknown event induces expression of the superantigen (systemically or just locally) which then activates V β 7-positive T cells. The cytokines produced by these T cells directly destroy pancreatic β -cells and/or upset the immunoregulatory balance, starting a local chain-reaction. c, Another possibility is that glucose intake, together with defective pancreas function, provokes hyperglycaemia, which in turn induces production of endogenous retrovirus transcripts in haematopoietic cells.

But, on activation, they rapidly invade the islets and destroy the β -cells.

Although this is certainly an attractive hypothesis, some of the predictions that could have been made do not seem to hold. For example, one might have expected a systemic expansion of V β 7-positive T cells and, perhaps, some restriction in the repertoire of V β 7-positive cells in the islets of recently diagnosed diabetics. But neither was observed in the two patients studied⁴, and alternative models need to be considered.

First, perhaps the retrovirus-encoded superantigen is not the initial trigger but rather a downstream precipitator. A completely unrelated event could stimulate the T cells to invade the pancreatic islets (Fig. 1b). The resulting inflammation, known as insulinitis, would remain innocuous until provoked to destruction by any of several possible secondary events, including a systemic or local response to a superantigen. Studies of non-obese diabetic mice have already shown that the progression from insulinitis to diabetes is neither automatic nor immediate, but that it is highly regulated⁸. Second, expression of the retroviral genome and the superantigen that it encodes may be a marker of, rather than the trigger for, diabetes (Fig. 1c). In support of this theory, endogenous retrovirus transcripts can be induced by the activation of leukocyte subsets under diverse conditions⁹, and even by the elevated glucose

levels found in diabetic mice¹⁰.

By identifying an endogenous retrovirus-encoded superantigen in man, Conrad *et al.* have made an exciting advance. We now need to define its precise relationship to diabetes, and to know whether it is generally linked to this disorder or whether it is found only in particular variants, such as patients with the more acute forms. It will also be interesting to see whether endogenous retroviral superantigens operate similarly in other autoimmune diseases. There is already evidence that an endogenous retrovirus is found specifically in patients with multiple sclerosis¹¹ but, again, it is critical to establish whether the retrovirus is the initial trigger for, a downstream precipitator of, or just a marker for, this disease. □

Christophe Benoist and Diane Mathis are at the Institut de Génétique et de Biologie Moléculaire et Cellulaire (CNRS/INSERM/ULP), 1 rue Laurent Fries, 67404 Illkirch, C.U. de Strasbourg, France.

1. Bach, J. F. *Endocr. Revs* 15, 516–542 (1994).
2. Tisch, R. & McDevitt, H. *Cell* 85, 291–297 (1996).
3. Conrad, B. *et al. Cell* 90, 303–313 (1997).
4. Conrad, B. *et al. Nature* 371, 351–355 (1994).
5. MacDonald, H. R. & Acha-Orbea, H. *Nature* 371, 283–284 (1994).
6. Gaskins, H. R., Prochazka, M., Hamaguchi, K., Serreze, D. V. & Leiter, E. H. *J. Clin. Invest.* 90, 2220–2227 (1992).
7. Suenaga, K. & Yoon, J.-W. *Diabetes* 37, 1722–1726 (1988).
8. André, L. *et al. Proc. Natl Acad. Sci. USA* 93, 2260–2263 (1996).
9. Moroni, C. & Schumann, G. *Nature* 254, 60–61 (1997).
10. Leiter, E. H., Fewell, J. W. & Kuff, E. L. *J. Exp. Med.* 163, 87–100 (1986).
11. Perron, H. *et al. Proc. Natl Acad. Sci. USA* 94, 7583–7588 (1997).

Protein–protein interactions

Calcium turns turquoise into gold

Tullio Pozzan

Over the past few years, targeting of appropriately engineered and differently coloured green fluorescent protein (GFP) mutants has become popular for investigating many biological processes. These range from gene expression¹ to protein sorting², and from organelle shape and mobility^{3,4} to diffusion of proteins within membranes or the lumens of organelles^{4,5}. But it has been difficult to make GFP sensitive to its local chemical environment — for example, to fluctuating concentrations of important ions and messengers such as Ca²⁺, H⁺ and cyclic AMP.

Advances in the molecular engineering of a GFP-based Ca²⁺ indicator are now reported by Anthony Persechini's group in the *Journal of Biological Chemistry*⁶ and *Cell Calcium*⁷, and by Miyawaki *et al.* on page 882 of this issue⁸. Taking advantage of fluorescence resonance energy transfer (FRET) between differently coloured GFP mutants, new, molecularly engineered Ca²⁺ indicators have been generated. These proteins can not only be targeted to different cellular organelles —

and so solve classic problems in Ca²⁺ signalling — but they also represent prototypes of new probes to investigate protein–protein interactions in living cells.

The unique ability of GFP to illuminate the behaviour of proteins in living cells was first shown by Chalfie and colleagues¹. Using an approach pioneered by Heim and Tsien⁹, GFP fusion proteins — dubbed 'cameleons' by Miyawaki *et al.*⁸ — have now been generated (Fig. 1). The two groups^{6–8} have linked blue-shifted mutants of GFP (emitting blue or turquoise light) to longer-wavelength GFP mutants (emitting green or yellow light), through two types of connecting sequence. FRET takes place between the two GFP mutants if they are appropriately positioned. A non-destructive, spectroscopic phenomenon, FRET is highly sensitive to the distance and orientation of fluorophores. The excitation of the donor results in nonradiative excitation of the acceptor, which eventually emits its characteristic fluorescence. In the cameleons⁸, when FRET is maximal, yellow (gold) light is emitted.

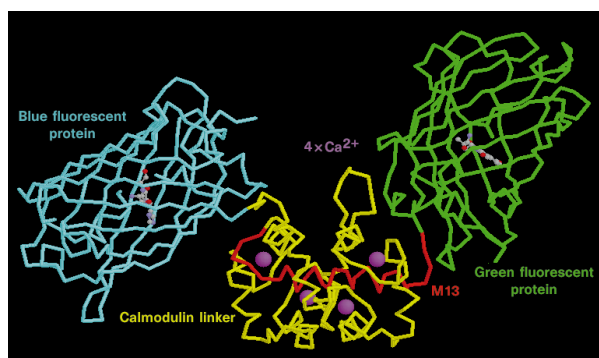


Figure 1 Romoser *et al.*⁶, Persechini *et al.*⁷ and Miyawaki *et al.*⁸ have linked blue-shifted GFP mutants to longer-wavelength GFP mutants through connecting sequences to which Ca^{2+} -calmodulin⁶ or Ca^{2+} alone^{7,8} can bind. The two GFPs are X-ray structures from J. Remington; the calmodulin-binding domain is the NMR structure from M. Ikura.

These chimaeric proteins can be used as Ca^{2+} indicators because a change in the concentration of Ca^{2+} results in a change in the ratio between the intensity of light that is emitted at two wavelengths — 510/440 nm in the studies by Persechini and colleagues^{6,7}, and 535/480 nm in the paper by Miyawaki *et al.*⁸. The short calmodulin-binding peptide used by Romoser *et al.*⁶ holds the two flanking GFPs at an optimal distance for efficient FRET in the absence of Ca^{2+} . In the work by Persechini *et al.*⁷, the amino- and carboxy-terminal EF hand domains of calmodulin have been introduced into the construct. Binding of unlabelled Ca^{2+} -calmodulin (or Ca^{2+} in ref. 7) straightens the linker peptide and decreases FRET.

In Miyawaki and colleagues' *cameleon* constructs, the linker between the two GFPs is calmodulin, fused to a calmodulin-binding peptide (M13). On binding Ca^{2+} , the calmodulin wraps around its adjacent target peptide, bringing the two GFPs closer to each other and increasing FRET. By building the calmodulin into the indicator⁸ or including calmodulin Ca^{2+} -binding domains⁷, the expression levels of the calmodulin and its target peptide are guaranteed to be equal. Moreover, the affinity and kinetics of Ca^{2+} -binding will be independent of the concentration of indicator; the likelihood of disrupting endogenous calmodulin-dependent signalling is reduced; and mutation of the calmodulin can tune the Ca^{2+} -sensitivity, to cover free Ca^{2+} from nanomolar to millimolar concentrations.

With the help of humanized codons and mutations to increase the level of expression and folding, the indicators can be introduced into cells by transfecting the gene⁸ rather than by microinjecting the protein^{6,7}. Transfection allows targeting sequences to be included, which send the indicator to specific compartments such as the nucleoplasm and the lumen of the endoplasmic reticulum. Targeted *cameleons* already seem to be superior to the other tools that are available for monitoring the dynamics of Ca^{2+} in such compartments. For example, although the Ca^{2+} -sensitive photoprotein aequorin can be targeted to cellular organelles with similar selectivity¹⁰, it requires the coenzyme coelenterazine, and its Ca^{2+} -dependent chemiluminescent signal is less amenable to single-cell imaging.

From the point of view of cellular Ca^{2+} homeostasis, the most important observation made by Miyawaki *et al.*⁸ is that the concentration of Ca^{2+} within the endoplasmic reticulum at steady state — an issue that has been widely debated during the past few years — is two to three orders of magnitude higher than in the cytosol. The authors report that the concentration of Ca^{2+} in their cell population ranged from around 400 μM to about 50 μM . The higher values were found in most cells, and were close to those measured with targeted aequorin and synthetic coelenterazine¹¹. The lower values were found in only about 10 per cent of cells (A. Miyawaki and J. Llopis, personal communication).

A final, highly interesting observation deserves a mention. One of Miyawaki and colleagues' constructs can monitor intermolecular rather than intramolecular FRET. From a general point of view, this result is the most exciting. Many groups, including mine, are pursuing the possibility of studying protein-protein interactions in living cells, using different protein partners tagged with GFP mutants and FRET to monitor their reversible association. To my knowledge, this is the first demonstration that a dynamically modulated protein-protein interaction can be followed using molecularly engineered GFP fusion proteins in live cells. Given the vast number of key cell functions that are governed by association between polypeptide partners, it is easy to predict that the new findings will induce many new laboratories to undertake this approach. □

Tullio Pozzan is in the Department of Biomedical Sciences, CNR Centre of Biomembranes, University of Padova, Via Colombo 3, 35121 Padova, Italy. e-mail: pozzan@civ.bio.unipd.it

- Chalfie, M., Tu, Y., Euskirchen, G., Ward, W. W. & Prasher, D. C. *Science* **263**, 802–805 (1994).
- Kaether, C. & Gerdes, H. H. *FEBS Lett.* **369**, 267–271 (1995).
- Rizzuto, R., Brini, M., Pizzo, P., Murgia, M. & Pozzan, T. *Curr. Biol.* **5**, 635–642 (1995).
- Cole, N. B. *et al. Science* **273**, 797–801 (1996).
- Subramanian, K. & Meyer, T. *Cell* **89**, 963–971 (1997).
- Romoser, V., Hinkle, P. M. & Persechini, A. *J. Biol. Chem.* **272**, 13270–13274 (1997).
- Persechini, A., Lynch, J. A. & Romoser, V. *Cell Calcium* (in the press).
- Miyawaki, A. *et al. Nature* **388**, 882–887 (1997).
- Heim, R. & Tsien, R. Y. *Curr. Biol.* **6**, 178–182 (1996).
- Pozzan, T., Rizzuto, R., Volpe, P. & Meldolesi, J. *Physiol. Rev.* **74**, 595–636 (1994).
- Montero, M. & Alvarez, J. *FASEB J.* (in the press).

Daedalus

A biased solid

Woven cloth is the oldest fibre-reinforced material. It is merely held together by the friction of the fibres threaded past each other, yet is amazingly durable. Furthermore, it has the useful property of being extremely weak in shear. A square of cloth can easily be pulled sideways into a parallelogram; instead of crossing each other at right angles, the warp and weft fibres merely do so at a slightly different angle. In couture, this property is neatly exploited in the 'bias cut', invented by Mme Vionnet in 1922. In a bias-cut garment the fibres run, not vertically and horizontally, but at 45°. It can stretch vertically, while shrinking horizontally to emphasize the wearer's figure.

Daedalus is now extending this idea to three dimensions. Just as two sets of fibres can be interwoven at right angles to make a flat cloth, so three sets at right angles could create a fibrous solid. A three-dimensional loom is as simple in principle, though more complicated in practice, than the normal planar variety. Daedalus's design has a square array of parallel threads running vertically, and two sets of shuttles at right angles running horizontally through the array. As each layer of interwoven horizontal threads is complete, beaters pack it down on the previous layers. The result is a solid block of three interwoven orthogonal sets of fibres.

This novel product will have many uses. Its toughness and tenacity, its ability to give slightly in any direction without breaking or even weakening, will make it ideal for all sorts of buffers, seals, shock-absorbers and load-spreaders. But Daedalus wants to cut it on the bias. The resulting 'Vionnet solid' would have enormous tensile strength, but very low shear strength. Push any two sides together, and they will come; but the other sides will extend to maintain its constant volume.

A shaped Vionnet solid will be a wonderfully forgiving engineering component. If it does not fit in the hole made for it, just stretch it till it does; once it is in place, just push it in and it will swell to grip its surroundings. In cotton, polyester or even metal wire, Vionnet blocks and forms will bring a new flexibility to construction and engineering. They will forgive amateurs their mistakes and relax the worries of professionals. Furthermore, when finally in place, they could be soaked in polymer resin, when they would harden into the first truly isotropic fibre-reinforced composite.

David Jones

Obituary

Hubert Horace Lamb (1913–97)

Climatologist who pioneered the study of climate change

Given the worries about the threat to climate posed by anthropogenic emissions of greenhouse gases, it is all too easy to forget that, until the 1970s, the meteorological establishment refused to recognize the inconstancy of climate on timescales of decades and centuries. For much of the twentieth century, atmospheric scientists considered climatology to be a rather dull, book-keeping exercise, and the study of climate change a fringe science.

Hubert 'H. H.' Lamb, who died on 28 June aged 83, did more than any other modern-day climatologist to change this state of affairs — he was among the first to alert the scientific community to the natural vagaries of climate over recent centuries and millennia, and to point to their possible effects on human societies.

Lamb came from a family that already had a distinguished scientific record. His grandfather, Horace Lamb, was a famous mathematician, author of textbooks widely used by meteorologists. His father was a professor of engineering. One aunt was a tutor at Newnham College, another lectured in archaeology. The young Lamb favoured history and languages at Oundle school, Northamptonshire, but was forced to abandon these subjects in favour of science because, as he later described it, of the "terrible concentration of expectations" placed upon him by his father.

But he rebelled. Starting his undergraduate career in natural sciences at the University of Cambridge, he defied his father and ended in geography. This experience of a catholic range of subjects at the secondary and tertiary levels of education provided a strong underpinning for the interdisciplinary nature of his later work.

Rebellion in the face of parental pressure, later to evolve into a propensity to question conventional scientific wisdom, became allied to a strong belief in the social responsibility of science and of individual scientists. This belief was fostered by discussions with Trevor Huddleston, a family friend who was later to become a prominent critic of apartheid, and Lewis Fry Richardson, a Quaker and a pioneer of numerical weather forecasting.

Lamb's career with the Meteorological Office, which he joined in 1936, almost came to an early end when he was asked to



tender his resignation having refused to accept an instruction to work on the meteorology of gas spraying during the war years. In the event, he was transferred to the Irish Meteorological Office where his main responsibility was forecasting for the new trans-Atlantic passenger flights. In 1945, Lamb resigned from this post after disagreeing with the director over the effect of increasing workloads on aircraft safety, and rejoined the British Meteorological Office the following year.

Working with limited weather data in Ireland during the war (when neutral Ireland was not allowed access to observations made from the United Kingdom) stood him in good stead during the late 1940s when he headed south to the Antarctic as expedition meteorologist on the whaler *Balaena*. It was during this time that Lamb, on the basis of his own observations in southern waters, began to doubt that climate was quite as constant as generally believed. His transfer to the climatology department of the Meteorological Office in 1954 proved a turning point. Lamb now began to establish an international reputation as a scientist able to cross disciplinary boundaries freely, making use of fragmentary data to weave a rich story of climate change and its effects on human affairs.

During the 1950s and 1960s, increasing contacts with interdisciplinary pioneers in the fields of history, botany and geology, amongst others, provided a basis for his most significant contribution to modern climatology, his two-volume treatise, *Climate: Present, Past and Future* (Methuen), which eventually appeared during the 1970s. Taking advantage of the extensive archives at the Meteorological Office, Lamb also produced a series of

shorter works on past variations in climate and the atmospheric circulation, their causes and effects on humanity, that have provided a lasting inspiration to later generations of researchers. It is, perhaps, as a source of modest and unassuming inspiration to scholars of climatic variation that H. H. Lamb will be best remembered.

In 1963, Lamb had been awarded a special merit promotion in recognition of his work on Antarctica and thus had the freedom to pursue his own research interests. But by the end of the 1960s even that arrangement proved insufficient to support his research ambitions. In 1971, he left the Meteorological Office and established the Climatic Research Unit at the University of East Anglia as a base for interdisciplinary work on climate change. He was director of the unit for six years, before retiring, and again broke new ground by creating a research group entirely funded by external grants and contracts; now, a quarter of a century later, the unit is recognized as a world authority not only on natural climate variability but also anthropogenic change. During this period, he also established a public presence, becoming the person the media turned to during flood or drought, heatwave or blizzard, thereby establishing a broader awareness of the vagaries of climate outside the scientific community.

After retirement, Lamb published two editions of his highly regarded account of climate and human affairs, *Climate, History and the Modern World*, also with Methuen. He was awarded the Symons Gold Medal by the Royal Meteorological Society in 1987 and completed his autobiography shortly before his death.

During his later years, Lamb was sceptical of certain claims regarding the dangers posed by global warming. An empiricist at heart, and well aware of the complexities of the climate system, he felt that climate models were limited in their ability to provide accurate forecasts. As he observed in 1994, "there has been too much theory and not enough fact in predicting the future". He had found a new orthodoxy to challenge.

Yet Lamb, as much as any scientist, prepared the ground for the research on anthropogenic climate change, and the public and political acceptance of the threat, that would, in 1992, result in the United Nations Framework Convention on Climate Change — the international community's initial response to this pressing environmental problem.

Mick Kelly

Mick Kelly is in the Climatic Research Unit, University of East Anglia, Norwich NR4 7TJ, UK. e-mail: m.kelly@uea.ac.uk

Sources of the Maelstrom

The Lofoten Maelstrom on the northern coast of Norway has been renowned for centuries for its strength and dangerous whirlpools. We now complete a previous review¹ of the historic literature about the Maelstrom and present results of simulations describing the large-scale dynamics of this remarkable phenomenon.

The Maelstrom is located at 67° 48' N, 12° 50' E between Lofoten Point (Lofotodden) and the island Værøy southwest of the main chain of the Lofoten Islands, and takes its name, Moskstraumen, from the small island Mosken in the centre (Figs 1, 2). The current is of tidal origin, although the prevailing northward currents and southwesterly wind may contribute to its strength. It is said to run at a speed of up to 5 or 6 m s⁻¹, but no current records are available for estimates of these extremes. Eddies and current shear-zones appear on satellite images².

The earliest accounts of the Maelstrom survive in old Nordic tales³. They tell of a pair of ponderous and magical millstones which sank either near the Pentland Firth north of Scotland or on the northwestern coast of Norway. Their continuous grinding of salt is said to form large eddies in the sea above. The term maelstrom is commonly thought to derive from the Dutch verb *malen* (Nordic *male*) meaning 'to grind'.

The first written account is probably by Olaus Magnus (1490–1558), a Swedish bishop in Rome⁴. On *Charta Marina* (1539) he drew a large eddy west of Lofoten, which he described as stronger than the Sicilian Charybdis, and attributed it to a divine force. Subsequently, fictional descriptions of a huge eddy appeared, were reprinted, and cited in mainstream seventeenth and eighteenth century European geographical literature¹. Contemporary Norwegian authors gave more factual descriptions^{5–8}.

Petter Dass (1647–1707), a Norwegian priest, produced a detailed and realistic description of Moskstraumen, as well as other strong tidal currents, in his poem, *The Trumpet of Nordland*⁶, written around 1685. He attributed its strength to the phases of the Moon — the current being strongest at full and new moons and weakest at half moons. Unfortunately, Dass was not translated into English. A report by Schelderup⁷ (written before 1751), noted the rotation of the current during the tidal cycle and attributed the driving force to differences in sea level across the Lofoten islands.

The celebrated stories by Edgar Allan Poe (*A Descent into the Maelström*, 1841) and Jules Verne (*Vingt Mille Lieues sous la Mer*, 1869), featuring the Lofoten Charybdis, rely heavily on Nordic sources. Poe's detailed

knowledge of the landscape and local tales, can be traced back to a fictitious description by Jonas Ramus from about 1715, which includes elements from an older source⁵. Ramus's description was quoted by Pontoppidan⁸ and reprinted in the sixth edition of *Encyclopaedia Britannica* (1823). Poe also acquired some information on local fishing practice, probably through the Anglo-American sources mentioned by Mabbott⁹.

No substantial modern studies of this strong tidal current have been reported. The Norwegian pilot book¹⁰ provides some information albeit of limited scientific value. We have therefore developed a high-resolution depth-integrated tidal model with a 0.5 km grid size (Fig. 1). This has enabled a study of the transition of the tide from a northwards progressive wave¹¹ on the outer shelf to standing oscillations in the fjords, and the enhancement of the tidal current particularly around Lofotodden. Input on lateral boundaries (sea level and volume flux) were obtained by interpolation from coarser models^{11,12}.

The simulated amplitude and phase for the dominant semi-diurnal tidal component M₂ (Fig. 1) agree well with observations. A strong sea-level gradient appears across the island chain with up to 25 cm higher amplitudes in Vestfjorden inside the islands than outside. This drives the current around Lofotodden and through the narrow channels between the islands further east. The sea-level variation, with contour lines converging on Lofotodden (Fig. 1) is due to the change in shelf width from a relatively broad shelf south of Lofoten to a narrow shelf further north, and the scatter-

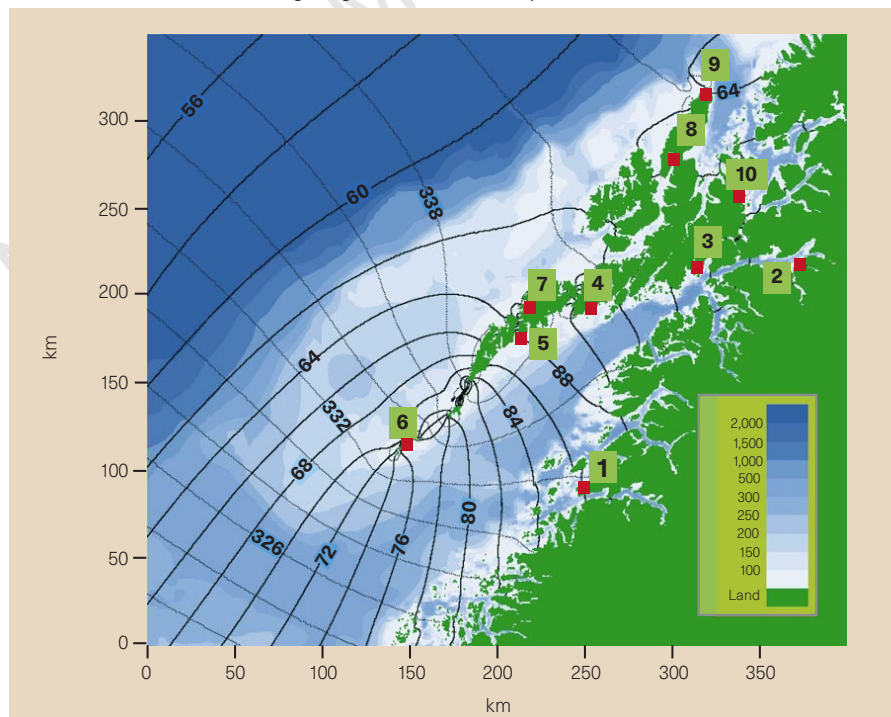


Figure 1 Modelled sea-level amplitude and phase for the dominant semi-diurnal tidal component (M₂). The Lofoten Islands extend northeast from Røst (6), to Lødingen (3). Vestfjorden between Lofoten and Bodø (1), Narvik (2), Andenes (9) and Harstad (10). Colour depth-scale is in metres. Solid contour lines show amplitude (in cm) with 2-cm separation. Dotted contour lines indicate phase (deg, GMT) with 2 deg separation. A decrease of 2 deg in phase corresponds to a time delay of the tide of 4 min. Observed/modelled amplitude (in cm) for the stations 1–10: 87/86, 99/100, 93/96, 93/92, 88/89, 78/76, 62/63, 68/66, 65/63 and 69/66, respectively.

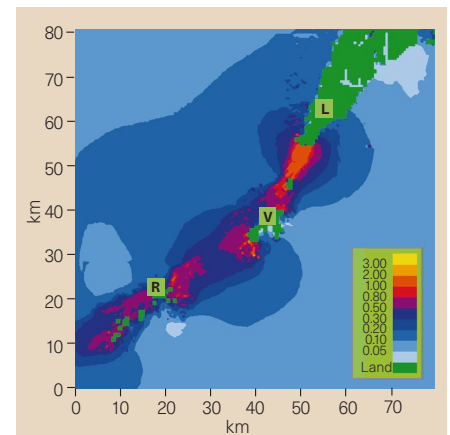


Figure 2 Maximum depth mean current for M₂ in the area southwest of Lofotodden (L) with Værøy (V) and Røst (R). Mosken is midway between L and V. Colour scale is in m s⁻¹. Volume flux through the sections between L and V and between V and R are 0.35×10^6 m³ s⁻¹ and 0.55×10^6 m³ s⁻¹, respectively, at the peak of the tide. Volume fluxes at spring tide are 1.8 times larger.

ing of the northbound tidal wave by the islands and shallow bank to the southwest.

The mean currents are predicted to be of the order of 0.1 m s^{-1} in deep water and up to 3.0 m s^{-1} on the narrow ridge (50–100 m deep) between Lofotodden and Røst (Fig. 2). A lack of data from this area hampers validation of the model; however, we have made comparisons with measurements from channels between the islands east of Lofotodden at Napp, Sundklakk and Gimsøy. Current shifts are predicted to occur about 2 h before high water, with a northwards current at high water and a southwards current at low water, in agreement with observations¹⁰.

A weak, roughly 6-km diameter, clockwise eddy appears in the simulations, centred about 5 km southwest of Lofotodden at the time of current shift on a rising sea. A similar sized anticlockwise eddy appears nearer Lofotodden at the time of current shift on falling sea. The current speed in these eddies is roughly 0.1 m s^{-1} , no comparison to the eddy in ancient literature. The eddies owe their existence to sea-floor topography and friction. West and east of Mosken the current vector rotates clockwise in nearly circular ellipses which may have been interpreted as a large eddy by early observers.

The semi-diurnal components, S_2 and N_2 , show a similar amplitude pattern to M_2 . The diurnal component, K_1 , which interacts with shelf wave modes, produces dominant diurnal currents in Sortlandsundet between stations 7 and 8 (Fig. 1), where the M_2 current is weak, matching observations (Norwegian Hydrographic Service, 1994). The generation and advection of small-scale eddies in the strong tidal jet remain a subject for future studies but these local effects are not likely to alter the large-scale patterns of sea-level variations reported here.

B. Gjevik, H. Moe

A. Ommundsen

Department of Mathematics, University of Oslo,
0316 Blindern, Oslo, Norway
e-mail: bjornng@math.uio.no

- Peterson, R. G., Stramma, L. & Kortum, G. *Progr. Oceanogr.* **37**, 1–115 (1996).
- Wahl, T. *Nordic Space Activities* 2–3, 22–23 (1995).
- Guerber, H. A. *The Norsemen* 128–130 (Senate Studio Editions, London, 1994).
- Magnus, O. *Historia om de Nordiska Folken* Swedish edn, 89–90 (Gidlunds, Stockholm, 1982; first published Rome, 1555).
- Schønnebøl, E. H. in *Historisk-Topografiske Skrifter om Norge og Norske Landsdele* (ed. Storm, G.) 178–218 (A. W. Brøgers Bogtrykkeri, Christiania, Oslo, 1895; original from c. 1590).
- Dass, P. *Nordlands Trompet* (ed. Seip, D. A.) 94–98 (Aschehoug, Oslo, 1947).
- Schelderup, A. *Kongelige Norske Videnskabskabers Skrifter* **2**, 1, 78–85 (Trondheim, Norway, 1824).
- Pontoppidan, E. *Norges Naturlige Historie* 123–140 (Reprint by Rosenkilde og Bagger, Copenhagen, 1977; first printed Copenhagen, 1752; English edn, London, 1755).
- Mabbot T. O. *Collected works of Edgar Allan Poe* II, 574–597 (Belknap, Cambridge, MA, 1978).
- Den Norske Los* 5, 6th edn, 241–242 (Norwegian Hydrographic Serv., Stavanger, 1987).
- Gjevik, B. in *Modeling Marine Systems* (ed. Davies, A. M.) **1**, 187–219 (CRC Press Inc, Boca Raton, Florida, 1990).
- Gjevik, B., Nøst E. & Straume T. J. *J. Geophys. Res.* **99**, C2, 3337–3350 (1994).

Morbillivirus in monk seal mass mortality

We have identified a morbillivirus in organs of Mediterranean monk seals (*Monachus monachus*), lost during a recent mass die-off. About half of the population of this highly endangered species, which inhabits the Mauritanian coast of West Africa (Cap Blanc), were killed. The outbreak is reminiscent of several recent morbillivirus outbreaks in aquatic mammals.

In 1988 there was a mass die-off of the two indigenous seal species of northwestern Europe, eventually leading to the deaths of about 20,000 animals. We and others identified a previously unrecognized morbillivirus, phocine distemper virus (PDV), as its primary cause. In subsequent years, several other mass mortalities of pinniped and cetacean species could be attributed to infections with known and newly identified morbilliviruses^{1,2}. This prompted us to speculate about the possible threat of morbillivirus infections to highly endangered pinnipeds like the Mediterranean monk seal (*Monachus monachus*)³. Two subpopulations of this species exist, one living off the Mediterranean coasts and the other off the Mauritanian coast of West Africa (Cap Blanc), with estimated population sizes of about 500 and 270 animals, respectively.

We and others have observed significant mortality in the West African population since May this year. After finding about 40 dead animals in and around the two caves that they predominantly inhabit, it became clear that a serious disease outbreak was ongoing, which eventually killed more than half of the total population. We autopsied several animals that had washed ashore during the outbreak. Most of these were in a state of advanced decomposition, probably because of the long period before washing ashore and the high ambient temperature. It was clear however, that most of these animals had suffered from respiratory distress, as emphysema and congestion of the lungs were among the most frequent macroscopic findings, consistent with previous morbillivirus infections of seals².

In an enzyme-linked immunosorbent assay³ for the detection of serum antibodies to canine distemper virus (CDV)¹, seven of 17 samples collected from the hearts of dead animals scored positive. We carried out virological analysis using organ samples from the 14 least decomposed carcasses. Using standard procedures² we isolated a cytopathic virus in Vero cells from lung and other organ samples of three of these animals. This virus also proved to be infectious for ferrets on parenteral infection. Our preliminary characterization of the virus with

polyclonal and monoclonal antibodies³ showed that the virus is a morbillivirus, probably different from CDV and PDV (data not shown).

We isolated total nucleic acids from lung samples with the guanidinium method⁴ and analysed them with a reverse transcriptase polymerase chain reaction (RT-PCR) using random hexanucleotides for first-strand synthesis. We used group-specific primers for the genus *Morbivirus*, which recognize the N gene (primer 1: 5'-ACAAACCA-NGGATTGCTGAAATGAT-3' and primer 2: 5'-CTGAAYTTGTTCTGAAYTGAGT-TCT-3') for amplification⁵. A semi-nested PCR using primer 2 and primer 3 (5'-ATC-GARACWATGTACCCGGC-3') increased the PCR signal for direct sequencing.

Sequencing allowed us to make a phylogenetic comparison with other morbilliviruses, using the Phylogeny Interference Package (provided by J. Felsenstein). We generated a maximum-likelihood tree, and assessed its stability using a bootstrap method (Fig. 1). We found the virus to be most closely related (mean distance, 0.030) to the previously described dolphin morbillivirus (DMV)². The antigenic difference between CDV and the isolated virus, the suboptimal state of the samples, and the fact that morbillivirus infections cause profound immunosuppression, may collectively explain why not all the dead monk seals tested had developed antibodies to CDV.

Further studies, preferably of live and freshly autopsied animals, are needed to confirm that this outbreak, reminiscent of many recent morbillivirus outbreaks in aquatic mammals, is indeed caused by infection with this newly identified monk seal morbillivirus (MSMV). In this light it is noteworthy that we could not detect the

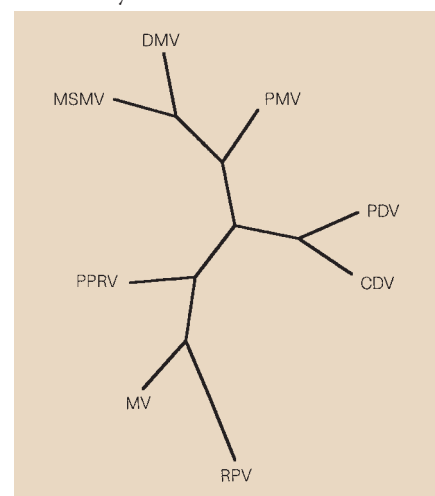


Figure 1 A maximum likelihood tree of N protein fragment sequences (121 nucleotides) of different morbilliviruses. DMV, dolphin morbillivirus; PMV, porpoise morbillivirus; PDV, phocine distemper virus; CDV, canine distemper virus; RPV, rinderpest virus; MV, measles virus; PPRV, peste des petits ruminants virus; MSMV, monk seal morbillivirus.

presence of paralytic shellfish poison in seal carcasses and in a composite mussel sample from the area. Further characterization of the virus as well as investigation of its origin are ongoing, and a rescue and rehabilitation programme for the remaining animals has been initiated by an international collaborative effort.

Albert Osterhaus, Jan Groen

Hubert Niesters

Erasmus University Hospital Rotterdam,
Institute of Virology, dr. Molewaterplein 50,
3015 GE Rotterdam, The Netherlands

Marco van de Bildt, Byron Martina

Lies Vedder

Seal Rehabilitation and Research Centre,
Hoofdstraat 94A, 9968 A6 Pieterburen,
The Netherlands

Joseph Vos, Hans van Egmond

National Institute of Public Health and The
Environment,

PO Box 1, 3720 BA Bilthoven, The Netherlands

Ba Abou Sidi, Mohamed Ely Ould Barham

Centre National de Recherches Oceanographiques et
des Peches,

B.P. 22, Nouadhibou, Mauritania

1. Osterhaus, A. D. M. E. & Vedder, E. J. *Nature* **335**, 20 (1988).
2. Barrett, T. et al. *Vet. Microbiol.* **44**, 261–265 (1995).
3. Osterhaus, A. D. M. E. et al. *Vet. Rec.* **130**, 141–142 (1992).
4. Chomczynski, P. & Sacchi, N. *Anal. Biochem.* **162**, 156–159 (1987).
5. Baron, M. D. & Barrett, T. J. *Gen. Virol.* **76**, 593–602 (1995).

α-Synuclein in Lewy bodies

Lewy bodies, a defining pathological characteristic of Parkinson's disease and dementia with Lewy bodies (DLB)^{1–4}, constitute the second most common nerve cell pathology, after the neurofibrillary lesions of Alzheimer's disease. Their formation may cause neurodegeneration, but their biochemical composition is unknown. Neurofilaments and ubiquitin are present^{5–8}, but it is unclear whether they are major components of the filamentous material of the Lewy body^{9,10}. Here we describe strong staining of Lewy bodies from idiopathic Parkinson's disease with antibodies for α-synuclein, a presynaptic protein of unknown function which is mutated in some familial cases of the disease¹¹. α-Synuclein may be the main component of the Lewy body in Parkinson's disease. We also show staining for α-synuclein of Lewy bodies from DLB, indicating that the Lewy bodies from these two diseases may have identical compositions.

We studied formalin- or ethanol-fixed, paraffin-embedded tissue sections of substantia nigra from six patients with idiopathic Parkinson's disease, and from four patients with DLB (all clinically and neuropathologically confirmed cases), as well as

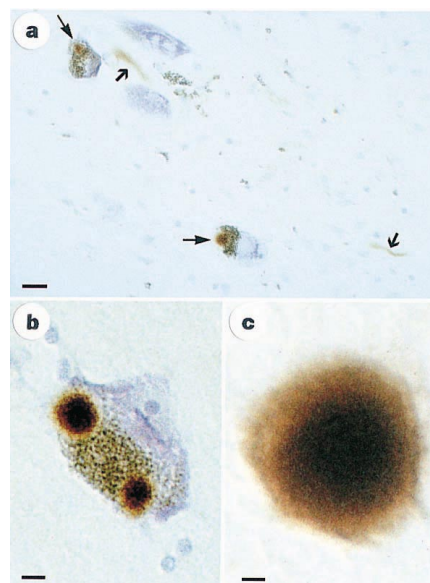


Figure 1 Substantia nigra from patients with Parkinson's disease (from the MRC Cambridge Brain Bank) immunostained for α-synuclein. **a**, Two pigmented nerve cells, each containing an α-synuclein-positive Lewy body (thin arrows). Lewy neurites (thick arrows) are also immunopositive. Scale bar, 20 μm. **b**, A pigmented nerve cell with two α-synuclein-positive Lewy bodies. Scale bar, 8 μm. **c**, α-Synuclein-positive, extracellular Lewy body. Scale bar, 4 μm.

cingulate cortex from the DLB patients (one with additional Alzheimer's disease pathology). We stained the tissue with affinity-purified, anti-α-synuclein serum PER2 (diluted 1:200)¹². This antibody, raised against a synthetic peptide corresponding to residues 116–131 of human α-synuclein, specifically recognizes α-synuclein on immunoblots of human cerebral cortex extracts and does not cross-react with the related β-synuclein¹². We also stained the sections with anti-β-synuclein serum PER3 (diluted 1:200) raised against a peptide corresponding to residues 99–111 of human β-synuclein¹². As an absorption control we

pre-incubated diluted PER2 overnight at 4 °C with 10 μM recombinant human α-synuclein. For immunohistochemistry we used avidin–biotin, with diaminobenzidine as the chromogen^{6,12}.

Substantia nigra sections from Parkinson's disease and DLB, and cingulate cortex sections from DLB and DLB with Alzheimer's disease incubated with the α-synuclein antibody PER2 showed staining of numerous brainstem-type and cortical Lewy bodies (Figs 1 and 2). Lewy neurites, which are dystrophic processes with the same immunohistochemical staining profile as Lewy bodies, were also reactive (Figs 1a, 2a). The strong staining made it difficult to distinguish between the core and the corona of the brainstem-type Lewy bodies (Figs 1, 2b). The staining was specific, and did not occur after pre-adsorption of the primary antibody with recombinant α-synuclein.

We obtained similar results with antibody PER1, raised against a synthetic peptide corresponding to residues 11–34 of human α-synuclein, indicating that full-length α-synuclein may be present in the Lewy body (data not shown). We found no specific staining of Lewy bodies or Lewy neurites with the β-synuclein antibody PER3.

We also stained tissue sections from Parkinson's disease and DLB with the ubiquitin monoclonal antibody 1510 (Chemicon, diluted 1:500)¹³. Double-staining of substantia nigra sections from Parkinson's disease with PER2 and 1510, showed staining of similar numbers of Lewy bodies and neurites with each antibody. Similarly, in adjacent tissue sections of substantia nigra and cingulate cortex from DLB, we found comparable numbers of Lewy bodies and neurites stained with PER2, antibody 1510 or undiluted neurofilament antibody RMO32, a monoclonal antibody to a phosphorylated epitope in the mid-sized neurofilament subunit which

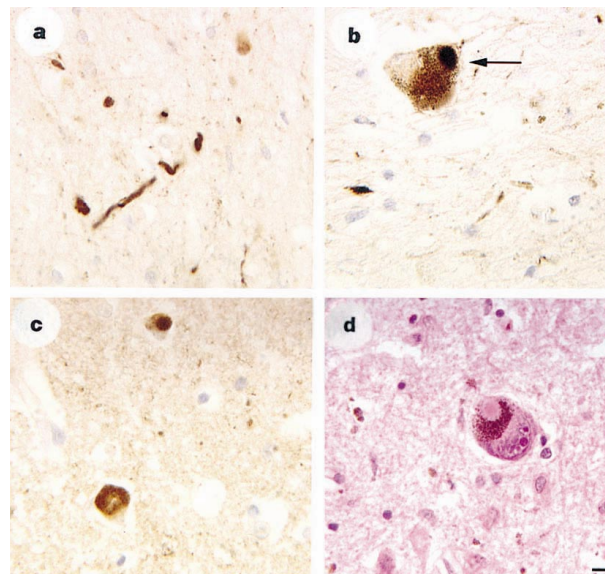


Figure 2 Tissue from patients with DLB (from the tissue collection of the Department of Pathology and Laboratory Medicine, University of Pennsylvania) immunostained for α-synuclein. **a**, α-Synuclein-positive Lewy neurites in the substantia nigra. **b**, α-Synuclein-positive Lewy body (arrow) in pigmented nerve cell of the substantia nigra. **c**, Two α-synuclein-positive Lewy bodies in the cingulate cortex. **d**, Haematoxylin and eosin-stained section of substantia nigra with a pigmented nerve cell containing a Lewy body. Scale bar for **a–d**, 10 μm.

specifically recognizes Lewy bodies⁶. Cortical Lewy bodies immunoreactive to PER2 had a similar morphology to the ubiquitin- and neurofilament-positive Lewy bodies in that they had round, oval or irregular shapes and frequently displaced the nucleus to one side (Fig. 2c).

The strong α -synuclein staining of brainstem-type and cortical Lewy bodies in idiopathic Parkinson's disease and DLB shows that α -synuclein is a component of the Lewy body. In some familial cases of Parkinson's disease there is an alanine to threonine mutation at residue 53 of α -synuclein¹¹. A major effect of this mutation may be to promote the aggregation of α -synuclein into filaments, resulting in the formation of Lewy bodies. The Parkinson's disease cases that we studied were non-familial, so at least two distinct pathogenetic mechanisms can lead to α -synuclein aggregation. α -Synuclein aggregation and Lewy-body formation may be important in the aetiology and pathogenesis of all cases of Parkinson's disease.

As brainstem-type and cortical Lewy bodies from DLB were strongly immunoreactive for α -synuclein, α -synuclein aggregation may also underlie Lewy-body formation in this condition. The intracytoplasmic Lewy body is therefore central to the neurodegenerative process, and both Parkinson's disease and DLB may be α -synuclein diseases.

Maria Grazia Spillantini

Medical Research Council Centre for Brain Repair and Department of Neurology, University of Cambridge, Robinson Way, Cambridge CB2 2PY, UK

Marie Luise Schmidt

Virginia M.-Y. Lee

John Q. Trojanowski

Department of Pathology and Laboratory Medicine, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104-4283, USA

Ross Jakes, Michel Goedert

Medical Research Council Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

Endogenous proviruses as "mementos"?

In a recent News and Views¹, J. P. Stoye discussed the potential health risks associated with the spread of endogenous proviruses from pigs and primates through the use of their organs for transplantation into humans. Although the risks to patients and the public from horizontal transmission may be as manageable as he represents, we would like to comment on his reference to the presence of vertically transmitted proviruses in the genomes of all mammals as "mementos".

The spread of such parasites throughout the genome of a species is expected to be accompanied by a large number of additional 'random' insertions. Many (perhaps most) of these insertions will be associated with mutations that are ultimately eliminated by natural selection, so the "mementos" are the surviving insertions, those with minimal effects on the host's fitness. For example, laboratory experiments show that the recent and rapid spread of the *P* element throughout the genomes of all *Drosophila melanogaster* in nature was accompanied by enormous transient net reductions in fertility and viability in the population^{2,3}. But in a few thousand *Drosophila* generations, there will be only "mementos" of the *P* element in the genomes of *D. melanogaster* (much as in seen in other species of *Drosophila*).

The apparently innocuous presence of these parasites does not mean that they have not caused a great toll in the past. Further, there is evidence that host genomes may have evolved the ability to repress rapid transposition of families of retroviruses⁴ or retrotransposable elements⁵, so that mutation of the relevant genes leads to element mobilization. There is therefore a risk that xenotransplantation could result in such mobilization in a foreign genetic background that lacks appropriate genes to repress transposition. Indeed, Stoye notes that mammalian hosts have evolved "adaptations of host viral-receptor proteins" that contain the spread of the endogenous proviruses. The ubiquitous presence of such effective adaptations implies a strong selective force (differential viability and/or fertility) on the hosts in the past.

In the absence of effective control of reproduction, there is a real risk that a germline infection in a xenotransplantation patient could introduce a new and potentially virulently mutagenic endogenous provirus into the human gene pool. Although the risks associated with insertional mutations are distributed over future generations, they are nevertheless great and worthy of serious consideration in the evaluation of risks and benefits inherent to an

expansion of xenotransplantation.

Charles H. Langley

Center for Population Biology, University of California, Davis, California 95616, USA

Brian Charlesworth

Institute of Cell, Animal and Population Biology, University of Edinburgh, Edinburgh EH9 3JT, UK

Stoye replies—On the basis of analogies between *P* elements and retroviruses, Langley and Charlesworth suggest that one potential risk associated with xenotransplantation is a form of insertional mutagenesis resulting from germline integrations by retroviruses derived from endogenous proviruses in the engrafted organs. I agree that these elements are potentially hazardous, but I am not convinced that the threat posed by this form of genomic bombardment is great enough to figure significantly in risk-benefit analysis of xenotransplantation. Rather, the much greater risk is that posed by these elements acting as infectious agents of disease.

Experiments in mice show that exogenous infection by retroviruses⁶, or activation of endogenous proviruses⁷, can result in germline infection. However, even under the most favourable circumstances, the number of acquired proviruses seems to be less than one per generation, of which at most one in ten is likely to be mutagenic^{8,9}. Increases in the number of germline proviruses will require further cycles of maternal expression followed by oocyte infection¹⁰, a process predicted to be significantly less efficient than *P*-element mobilization associated with hybrid dysgenesis in *Drosophila*.

To pose any significant risk to human fitness, a very high frequency of horizontal retroviral transmission (followed by infection of the germ line) would be required. For horizontal transmission on such a scale to occur, levels of viraemia associated with unacceptably high probabilities of virally induced disease would almost certainly have to occur. Several groups are currently addressing the question of whether xenotransplantation procedures pose any threat of infectious disease; provided that this concern is met, the threat to the human gene pool seems exceedingly remote.

Jonathan Stoye

National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK
e-mail: jstoye@nimr.mrc.ac.uk

- Lewy, F. H. in *Handbuch der Neurologie* Band III (ed. Lewandowski, M.) 920–933 (Springer, Berlin, 1912).
- Trétiakoff, M. C. Thesis, Univ. Paris (1919).
- Okazaki, H., Lipkin, L. E. & Aronson, S. M. *J. Neuropathol. Exp. Neurol.* **20**, 237–244 (1961).
- Kosaka, K., Oyanagi, S., Matsushita, M., Hori, A. & Iwase, S. *Acta Neuropathol.* **36**, 221–233 (1976).
- Goldman, J. E., Yen, S.-H., Chiu, F.-C. & Peress, N. S. *Science* **221**, 1082–1084 (1983).
- Schmidt, M. L. *et al. Am. J. Pathol.* **139**, 53–65 (1991).
- Kuzuhara, S., Mori, H., Izumiya, N., Yoshimura, M. & Ihara, Y. *Acta Neuropathol.* **75**, 345–353 (1988).
- Lennox, G., Lowe, J., Morrell, K., Landon, M. & Mayer, R. J. *J. Neurol. Neurosurg. Psychiatr.* **52**, 67–71 (1989).
- Duffy, P. E. & Tennyson, V. M. *J. Neuropathol. Exp. Neurol.* **24**, 398–414 (1965).
- Roy, S. & Wolman, L. *J. Pathol.* **99**, 39–44 (1969).
- Polymeropoulos, R. *et al. Science* **276**, 2045–2047 (1997).
- Jakes, R., Spillantini, M. G. & Goedert, M. *FEBS Lett.* **345**, 27–32 (1994).
- Shaw, G. & Chau, V. *Proc. Natl Acad. Sci. USA* **85**, 2854–2858 (1988).

- Stoye, J. P. *Nature* **386**, 126–127 (1997).
- Kidwell, M. G., Kimura, K. & Black, D. M. *Genetics* **86**, 815–828 (1988).
- Engels, W. R. *Genetics* **145**, 11–15 (1997).
- Best, S., Le Tissier, P., Towers, G. & Stoye, J. P. *Nature* **382**, 826–829 (1996).
- Pelissou, A. *et al. EMBO J.* **13**, 4401–4411 (1994).
- Panthier, J. J., Condamine, H. & Jacob, F. *Proc. Natl Acad. Sci. USA* **85**, 1156–1160 (1988).
- Rowe, W. P. & Kozak, C. A. *Proc. Natl Acad. Sci. USA* **77**, 4871–4874 (1980).
- Jenkins, N. A. & Copeland, N. G. *Cell* **43**, 811–819 (1985).
- Lock, L. F., Jenkins, N. A. & Copeland, N. G. *Curr. Top. Microbiol. Immunol.* **171**, 27–41 (1991).
- Lock, L. F. *et al. EMBO J.* **7**, 4169–4177 (1988).

Transforming the theatre of surgery

A Commotion in the Blood: Life, Death and the Immune System

by Stephen S. Hall
Holt: 1997. Pp. 457. \$36

Fred S. Rosen

At the beginning of this century, William B. Coley was one of the leading cancer surgeons in New York City. Born a Connecticut yankee and reared in impoverished circumstances with a stern Wesleyan morality, he eventually found his way as an undergraduate to Yale University and thence to the Harvard Medical School — an educative sequence that only infrequently leads to the anticipated result. He then went to New York City where he rapidly developed an interest in the treatment of malignancies.

He observed the complete regression of a cancer in a patient who had recently been infected with the skin disease erysipelas, and this prompted him to make culture filtrates of *Streptococcus pyogenes*, the microbe that causes erysipelas, for injection into other cancer patients; he subsequently added *Serratia marcescens* to the brew that came to be called Coley's toxin. There were sporadic and tantalizing successes with Coley's toxin.

Coley wisely befriended the Rockefellers and Huntingtons and effectively tapped into their charitable instincts. This, ultimately, was a critical factor in the creation of the vast biomedical emporium on the East River at 68th street in Manhattan. James Ewing, of the eponymous bone cancer, Ewing's sarcoma, and head of pathology at the Memorial Hospital, became a strong advocate of Röntgen's discovery of X-rays in the treatment of malignancies, and so became the nemesis of Coley's toxin. The monumental conflict between these giants of surgery and pathology is a fascinating story in the history of medicine and is well told in the best part of this book.

In any event, Coley's toxin was never standardized or subjected to proper scientific scrutiny and interest in it almost completely died with him in 1936. Coley was at heart a surgeon, a profession with a strong theatrical element that is basically in conflict with the painstaking details of data-gathering requisite for good science. Ultimately, proper scrutiny of Coley's toxin proved that the cancer-shrinking component was endotoxin or lipopolysaccharide in the bacterial filtrate.

Endotoxin is a potent vehicle for eliciting the synthesis and secretion of inflammatory molecules such as interferons, interleukins-2 and -12 and tumour necrosis factor- α . How each of these molecules was discovered and the exciting race to clone the genes encoding them are recounted in gripping journalistic style in this book. The cast of



Coley (centre) in December 1892, one month before the first use of Coley's toxin.

characters involved leads to the inevitable conclusion that science is not a gentlemanly profession in the latter half of the twentieth century.

And there now enters another surgeon, Dr Steven Rosenberg, who exploits the clinical use of these molecules to treat cancer patients and elevates the theatre of surgery to hitherto unknown heights. Like Coley, he is well educated and has tapped into the vast resources not of the robber barons but of the US taxpayers who support the National Institutes of Health. His personalized war against cancer arouses sympathy and alarm. Torn between compassion for the needy and very sick patients he is treating and "careerist entrepreneurship", he does not emerge from the book as a model of how to overcome our frustration with the very limited progress in cancer therapy in recent years.

It is unfortunate that these therapeutic manoeuvres from Coley to Rosenberg have been called immunotherapy. They are really phlogistic therapies, which incite inflammation in a nonspecific manner, and have at best to do with innate and not adaptive immunity. The work of Thierry Boon in Belgium, which is also detailed in this book, has recently been heading in the direction of immunotherapy, *proprement dit*, as he defines tumour-specific antigens that may evoke future therapeutic possibilities.

A number of small peccadilloes mar this book in the interchangeable use of toxin and vaccine and other loose word usage. The

Royal Experiment — the smallpox inoculation of the children of Frederick, Prince of Wales, which the royal physician, Sir Hans Sloane, described as "raising such a commotion in the blood" — did not occur in 1721, nor is the *New England Journal of Medicine* a publication of the Harvard Medical School or the Massachusetts General Hospital.

"Careerist entrepreneurship" was a choice term used in an upbeat review of this book by Roy Porter in *The New York Times*. *A Commotion in the Blood* is a wonderful book for lay readers who want to know more about current biomedical science. Unlike James Gleick's *Genius* or Dava Sobel's *Longitude*, two recent popular books for the lay reader that give no real scientific insight into their subject matter, *A Commotion in the Blood* is replete with scientific details recounted in very comprehensible prose. It is indeed a good read for the aficionados of *The New York Times*, and even of the *London Times*, as well as any surviving readers of the *Boston Evening Transcript*, so maligned by T. S. Eliot.

Twenty years ago a godson of mine developed a very malignant lymphosarcoma. The sage doctors gave him six months to live. He is now well, prosperous and thriving. He never received any immunotherapy. Some day we will understand these still mysterious medical events. □

Fred S. Rosen is at the Center for Blood Research, Warren Alpert Building, 200 Longwood Avenue, Boston, Massachusetts 02115, USA.

Research remedies

Complementary Medicine: A Research Perspective

by Charles Vincent and Adrian Furnham
Wiley: 1997. Pp. 305. £24.95, \$44.95

Iain Smith

There continues to be widespread interest in complementary medicine both from consumers and from their vocal providers. But how do these groups handle the frequent criticism that there is little robust evidence that their therapies actually work?

Additions to the blossoming literature that reiterate the same message — that the research evidence for clinical effectiveness is poor — are always welcome. This new book is especially valuable because it aims to educate the interested researcher or critical analyst about other research aspects of complementary medicine. It includes welcome introductory sections on why people use complementary medicine and the sociology of illness.

But the section on research methods for evaluating effectiveness misses two opportunities. The first is to develop the model best suited for evaluating effectiveness when the randomized controlled trial is deemed either inappropriate or impractical. Many believe

that a modification of the $n=1$ trial approach, beloved of psychologists, is one way forward. It is a shame that the arguments for such radical alternatives to the randomized clinical trial are not developed more fully, particularly as practitioners of complementary medicine need to be encouraged to embrace the need for research, appropriate research training and systematic high-quality, outcome-led, clinical audits.

The second missed opportunity is that the authors could have said more about appropriate outcome measures. These measures are important in the evaluation of most therapies, and basic points about their development and use should be made.

The book does not include all complementary therapies. Notable deliberate omissions include aromatherapy (the panacea for midwifery problems), hypnotherapy and counselling (the panacea for the 1990s). What is more, its focus on particular therapies does not always help in condition-based prescribing — that is, in deciding which therapies might be best for which condition. Even that approach, however, would probably still be contentious.

As with all new textbooks, this one is already out of date. For example, there is nothing on the use of St John's wort (hypericum) for depression; the summary table comparing orthodox and complementary

medicine does not take into account the policy drive for evidence-based medicine and cost-effectiveness; and there is no mention of the recently funded UK Medical Research Council trial on manipulative therapies for chronic back pain in primary care or of UK moves to form a central chiropractic register, similar to that being put in place for osteopaths.

It is always easy to point out acts of omission when reviewing a book; it is almost like saying "Well, if I had written it, then...". This book is easy to read, free of overt values and jargon, and sensibly referenced and laid out. It is a clear account of some of the key issues facing both researchers into complementary medicine and their friends and critics. Although probably lacking in depth for most seasoned observers, it is ideal for beginners or those with a casual interest in the field. □

Iain Smith is at the Nuffield Institute for Health, 71–75 Clarendon Road, Leeds LS2 9PL, UK.

Everyman physics

Giant Molecules: Here, There, and Everywhere...

by Alexander Yu Grosberg and Alexei R. Khokhlov

Academic Press: 1997. Pp. 314. \$39.95, £24.95

Edwin L. Thomas

Who would have thought a pair of theorists would produce a very readable and perceptive monograph on polymer physics? Yet this is exactly what Alexander Yu Grosberg and Alexei R. Khokhlov have done in this attractive book.

The authors take us through the fundamentals of macromolecules in an intuitive and entertaining way. They hold the reader's interest by including occasional banter on the relevance of the topic being discussed. They question whether it is sufficiently important to be included and whether their treatment of it will be clear to the reader. We are treated not only to good physics, but also to some history and poetry mixed in with anecdotes about prominent polymer scientists.

The topics covered include much of modern polymer statistical physics, especially polymer solutions and melts, as well as biopolymers and their role in molecular biology. Noteworthy chapters are the one on the "Physics of High Elasticity", which includes an excellent tutorial on the thermodynamic relationship of ideal rubber elasticity to ideal gases, and another on the "Dynamics of Polymer Fluids", which considers the critical role of entanglements and the concept of reptation in explaining the viscosity and diffusivity of macromolecules.

The mental pictures evoked of polymer chains and their mutual interactions give

It's a frog's life



This amorous pair of *Phyllomedusa trinitatis* is demonstrating one of the nine documented reproductive methods used by frogs native to the islands of Trinidad and Tobago. But not only is there plenty of variety in the sex life of these particular Caribbean frogs, the diversity of reptilian and amphibian life in general on these two islands is also remarkable. *Amphibians and Reptiles of Trinidad and Tobago* by John C. Murphy (Krieger, \$72.50, £66.50) is said to be the only book dealing with the entire herpetofauna of the islands, covering 130 species and subspecies. It includes identification keys, colour photographs and distribution maps.

deep insight into how these molecules behave in the dilute, semi-dilute and melt regimes. The authors go lightly with derivations, but the ones they do are easy to follow and well discussed in terms of physical consequences. The explanations of the interplay of entropy and internal energy in the coil-globule transition and phase transitions of polyelectrolyte gels are about the clearest I have read anywhere. The authors strongly believe that understanding the role of the primary, secondary and tertiary structure of biopolymers will become an increasingly important aspect of polymer physics. They put forward probability arguments to speculate about the role of chance in the origin of life and in the evolution of both the chemical and the biological properties of biomacromolecules.

Simple cartoons and colour plates depict the more spatially intriguing aspects of the various molecules and models. An accompanying CD-ROM (produced by S. Buldyrev of Boston University) greatly enhances the text, with simulations illustrating, for example, protein folding. What is more, the ambitious reader can alter the program parameters on the CD-ROM to see their influence and even produce customized movies. □

Edwin L. Thomas is in the Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

Oceans of truth

Cod: A Biography of the Fish that Changed the World

by Mark Kurlansky
Walker: 1997. Pp. 294. \$21

John Shepherd

If you want to know how a flush toilet may help in the cooking of dried or salted cod, what W. H. Auden and Louis MacNeice thought of the product, how it contributed to the causes of the American War of Independence, or why fishing for cod has been banned in Canadian waters for the past five years, you should read this book.

The belief that the oceans and their resources are limitless and invulnerable to harm from human activities persisted for a very long time. In about 400 BC, Clytemnestra asked: "There is the sea — who shall exhaust the sea?", in Aeschylus's drama, the *Oresteia*. In 1883, T. H. Huxley stated: "I believe, then, that the cod fishery... and probably all the great sea fisheries, are inexhaustible: that is to say that nothing we do seriously affects the number of fish. And any attempt to regulate these fisheries seems... to be useless." In the twentieth century, we have successfully disproved these hypotheses. The great whales were reduced almost to extinc-



The catch at Scarborough in 1932, when North Sea fish stocks were healthier.

tion. In two world wars, when fishing ceased in the North Sea for several years, the stocks recovered dramatically, demonstrating both that fishing did affect the number of fish and that regulation could be effective. Moreover, excessive fishing since then has caused the collapse of stocks of herring off Norway, and of herring and mackerel in the North Sea, all previously measured in millions of tons. The most dramatic and notorious collapse has however been that of the cod stocks on the Grand Banks of Newfoundland.

The history of this once great fishery, and the fish that supported it, is the central theme of this fascinating book by Mark Kurlansky, once a fisherman himself. He follows the story from the 'discovery' of great shoals of fish, previously the secret of canny Basques, by Giovanni Caboto (alias John Cabot) in 1497, to the present day. In so doing, he ranges widely. The importance of dried and salted cod as a staple food, sustaining international trade from the Vikings, through the American Civil War, until Clarence Birdseye invented frozen fish, is vividly described. He relates its link to the slave trade, and the rise of the codfish aristocracy of Boston, and much else besides. The story is brought to life with amusing anecdotes, keen observations and telling quotations from conversations with fishermen, scientists and politicians from countries all around the North Atlantic. Rightly, he identifies no single culprit: most disasters occur when several things go wrong at once, and this seems to be no exception.

This is, however, much more than just a history of the cod fisheries. Nicely produced in a fashionably small format, the book contains many aptly chosen quotations, excellent illustrations and numerous recipes for preparing cod, especially the dried and salted varieties. Even if your opinion of the product, like mine, coincides with that of Auden

and MacNeice, there is much to savour here.

With the latest research (*Nature* **385**, 521; 1997) suggesting that the cod stock of the North Sea is teetering on the edge of collapse, this is, above all, a cautionary tale we would do well to heed. The oceans may be vast, and their resources great, but so is our capacity to harm them. □

John Shepherd is at the Southampton Oceanography Centre, University of Southampton, Empress Dock, Southampton SO14 3ZH, UK.

Flights and fights

Concorde and the Americans: International Politics of the Supersonic Transport

by Kenneth Owen
Smithsonian Institution Press: 1997. UK publisher Airlife. Pp. 232. \$35, £24.95

Paul Duffy

The story of civil aviation and aircraft is usually told in terms of engineering, technical and marketing achievements, but equally interesting tales centre on the politics and economics of each important project.

This, according to the cover, is the story told in *Concorde and the Americans*. In fact, only two pages make any more than a passing reference to economics, and these, under the heading "uneconomy class", relate to figures from reports by British Airways and the former British Overseas Airways Corporation.

Concorde first took to the air in 1969 and has been flying daily services between New York and Europe for something over 20 years. But the development of supersonic transport (SST) was plagued by bitter political controversies over the environmental and economic effects and by strained relationships between the United

States, France and the United Kingdom. Kenneth Owen recounts the transatlantic negotiations that led to the rise and demise of plans for an American SST, to the Anglo-French collaboration in building Concorde and to the battle for approval of the Concorde service to the United States.

The book is essentially compiled from available government documents—minutes of cabinet meetings, departmental minutes, reports commissioned by government and industry from advisers and consultants—and from court records, mainly from US courts which heard the suits brought by groups trying to prevent Concorde's access to New York and the subsequent appeals.

It is therefore a useful reference book for future historians and researchers interested not only in the Concorde programme but also in aviation, technical and environmental questions that may arise in the future.

Owen covers all these aspects well, from both the UK and US sides, and gives full details of his sources. But, apart from minor quotations from joint British Aircraft Corporation/Aerospatiale press releases and brief reports of two conferences on future SST held in Toulouse in 1990 and 1994, there is almost no material originating from France. Surely Concorde was a joint Anglo-French programme?

Owen does, however, make clear the

difficulties of international programmes involving different bureaucracies, different programme targets and even different measurements—metric and imperial. Yet he outlines them only from a British point of view.

He shows that the French president and government were totally pro-Concorde; but that Harold Macmillan's government, thinking the French might back out of the programme after a few years, decided that no escape clause should be included in the two countries' Concorde treaty. This effectively locked Britain, but not France, into a technologically successful but financially disastrous and increasingly expensive programme.

We also learn that a main motivation for the Anglo-French programme was the belief that the United States was only a few years behind in developing its own SST. When environmental issues first infringed on aviation in the late 1960s and early 1970s, they became a critical factor in the cancellation of the US rival project. It quickly became evident that only the British and French national airlines would take and operate Concorde.

Regrettably, the author assumes that the reader is fully conversant with the environmental questions, and the huge political problems they raised for all technological matters at and since that time. Some background would have been useful.

International partnership was deemed necessary by Europe's aviation industry and its governments because of the high cost of developing the new technology for the next major step in civil aviation, which is how supersonic transport was seen in the 1950s and 1960s.

Unfortunately, and with hindsight, there were two inherent mistakes in this assumption: that the United States would be a partner when in fact it was not anxious to share in any joint project and felt it could go it alone; and that the immediate future of commercial air transport would be shaped by SST rather than high-capacity subsonic airliners—the 747, DC-10 and Tristar, built by the three giant US corporations Boeing, McDonnell Douglas and Lockheed. Europe delayed its entry into this important market by not seeing the signs early enough.

Indeed, the Bermuda Convention, where the United Kingdom and the United States agreed to recognize each other's airworthiness certification requirements, resulted in the approval of new US wide-body planes for British operation while Concorde ran into considerable difficulties in the United States.

Owen deals well with some of the environmental aspects. He points out that Concorde is still the only aircraft that has been required to undergo an environmental impact study. And he reveals how federal aviation requirements on noise and pollution were the focus for opponents trying to prevent Concorde flying to the United States.

He also shows that the US secretary for transportation, William T. Coleman, displayed a sound practical and judicial approach to settling the prolonged question of Concorde access, in part by leaving potential grounds for appeals for the courts to decide.

Coleman combined a sense of fairness towards the Concorde countries with an awareness of political reality by allowing a trial period of 16 months to determine whether approval would be long-term or not.

Owen concludes the book with a detailed look at future SST, concluding that larger subsonic aircraft will come first. He foresees considerable investment initially in 1,000-passenger aircraft for main routes rather than a 300-seater Mach 2/Mach 3 plane. And he queries the demand for a "son of Concorde" from the world's airlines.

He makes some mention of the other existing SST, the Soviet Tu-144. Sadly, the (mostly inaccurate) remarks about this reflect perceived Western opinion rather than the actuality. □

Paul Duffy is at 9 Deilginis, Shannon, County Clare, Ireland.



Bumpy ride ahead: Concorde makes one of its earliest public appearances, in 1968.

HUTTON GETTY IMAGES

Design and synthesis of chromophores and polymers for electro-optic and photorefractive applications

Seth R. Marder, Bernard Kippelen, Alex K.-Y. Jen & Nasser Peyghambarian

The ability of nonlinear optical materials to transmit, process and store information forms the basis of emerging optoelectronic and photonic technologies. Organic chromophore-containing polymers, in which the refractive index can be controlled by light or an electric field, are expected to play an important role.

Intense light, such as a laser beam, can change the optical properties of a nonlinear optical (NLO) material, and these changes will in turn affect the properties of the light beam as it propagates through the material. The study of the interaction of light with matter is the subject of nonlinear optics and embraces remarkable effects such as second-harmonic generation. For example, focusing an infrared laser pulse (wavelength 1,064 nm) into a potassium dihydrogen phosphate crystal can lead to a pulse with a green colour at the second harmonic (wavelength 532 nm). More generally, the NLO properties of materials can be used to control the phase, the state of polarization, or the frequency of light beams. They can be used to store and restore information optically, or to deflect light beams and so route optical information between fibre-optic channels. With the emergence of photonic technologies in areas such as telecommunications where information is coded, transported and routed optically, there is a strong technological demand for high-performance NLO materials^{1–3}. Second-order NLO materials look particularly promising for photonic applications because they exhibit an important effect called the linear electro-optic effect in which the refractive index of the material can be controlled through the application of an external electric field. This effect can be used to impress information on an optical carrier signal by modulating its phase or amplitude with an applied field varying in time and in amplitude. Such materials can therefore be used for the fabrication of ultra-fast integrated electro-optical modulators. When combined with photoconducting properties, this effect can be used to store information optically by modulating spatially the refractive index of a photorefractive material: such a material can be used as a data storage medium. To optimize the performance of these materials for photonic applications, a fundamental understanding of the interrelationship between their chemical and NLO properties is required.

Nonlinear optical effects arise from nonlinear polarization of molecules and materials. The ability of the charges in the material to be displaced by an electric field (such as that associated with light) is referred to as polarizability. In presence of low-intensity light, the induced polarization of a molecule is linearly proportional to the electric field strength of the light. However, in the presence of high-intensity light, the polarization of the molecule is no longer linear in the field. The induced polarization of the molecule is thus a nonlinear function of the field strength E , and can be approximated by a Taylor series expansion⁴:

$$P = \alpha E + (1/2!)\beta E^2 + (1/3!)\gamma E^3 \dots \quad (1)$$

where P is the induced polarization, α is the linear polarizability,

and β and γ are the first and second hyperpolarizabilities, respectively. Second-order nonlinear optical effects referred to above are a direct manifestation of nonlinear polarization arising through the $(1/2!)\beta E^2$ term of the Taylor series and only occur in molecules that do not have a centre of symmetry. Organic systems have several features that may make them important in information technologies based on NLO materials. In particular, they can show large nonlinearities, which have ultra-fast responses because the electronic polarization is nearly instantaneous. Furthermore, polymeric organic systems offer the possibility of systems that integrate photonics with existing silicon-based circuitry. Such integration would be difficult using inorganic nonlinear optical materials. But for the polymeric materials to exhibit bulk second-order optical nonlinearities, the non-centrosymmetric nonlinear molecules must be arranged in the materials in such a manner that the bulk material also does not have an effective centre of symmetry. We describe recent advances in the design of molecules with large second-order optical nonlinearities, and present strategies for improving the thermal stability of chromophores—an important practical consideration for real applications. Methods for incorporation of these dyes into electrically poled polymers (to achieve a non-centrosymmetric orientation of the molecules) with enhanced optical nonlinearity and temporal stability at elevated temperature are then discussed. The basic criteria for the observation of photorefractivity is then introduced, as is the evolution of the design strategies that have already led to high-performance photorefractive materials.

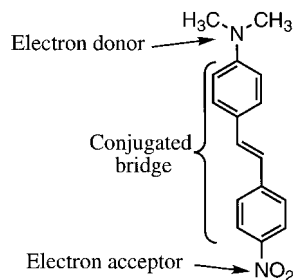
Design of molecules for nonlinear optics

In the early 1970s Oudar and Chemla suggested that a two-state model could be used to guide the design of second-order NLO chromophores⁵. In this model,

$$\beta \propto \left(\frac{\mu_{ge}^2(\mu_{ee} - \mu_{gg})}{E_{ge}^2} \right) \quad (2)$$

where μ and E are the dipole matrix element and transition energy, respectively, between the ground state (g) and the first strongly allowed charge-transfer excited state (e). Physically, the introduction of a $(\mu_{ee} - \mu_{gg})$ term meant that as the electrons interact with the oscillating electric field of light, they show a preference to shift from one direction along the axis of the molecule relative to the other. Accordingly, molecules for second-order NLO applications were based simply upon aromatic π -electron systems unsymmetrically end-capped with electron donating and accepting groups to impart the directional bias. A prototypical NLO chro-

mophore was 4-(*N,N*-dimethylamino)-4'-nitrostilbene (DANS), which is shown below.



In DANS, the two benzene rings and the double bond provide the conjugated π -system and the polarizable electrons, the dimethylamino group acts as the donor, and the nitro group acts as the acceptor. This model guided the design of second-order NLO chromophores for quite some time. However, a simple model has been proposed recently in which β is correlated with the degree of ground-state polarization^{6–8}. The degree of ground-state polarization, that is, the degree of charge separation in the ground state, depends primarily on the chemical structure (for example, the structure of the π -conjugated system, or the strength of the donor and acceptor substituents), but also depends on the surroundings of the molecule (for example, the polarity of the medium). In donor–acceptor polyenes, this variable is related to a geometrical parameter, the bond-length alternation (BLA), which is defined as the average of the difference in length between adjacent carbon–carbon bonds in a polymethine ((CH)_n) chain. Polyenes have alternating double and single bonds (bond lengths equal to 1.34 Å and 1.45 Å, respectively) and thus show a high degree of BLA (+0.11 Å). To better understand this correlation, it is helpful to discuss the wavefunction of the ground state in terms of a linear combination of the two limiting resonance structures: (1) a neutral form characterized by a positive BLA, and (2) a charge-separated form characterized by a negative BLA (as the double and single bond pattern is now reversed relative to the neutral form) (Fig. 1). For substituted polyenes with weak donors and acceptors, the neutral resonance form dominates the ground-state wavefunction, and the molecule has a high degree of BLA (Fig. 1a, b). With stronger donors and acceptors, the contribution of the charge-separated resonance form to the ground state increases, and BLA decreases. When the two resonance forms contribute equally to the ground-state struc-

ture, the molecule exhibits essentially no BLA. This zero BLA limit is the so-called cyanine limit, and refers to the common structure of a cyanine molecule. Such cyanine molecules are known to be correctly represented by two degenerate resonance forms, resulting in structures with virtually no BLA (Fig. 1d). Finally, if the charge-separated form dominates the ground-state wavefunction, the molecule acquires a reversed BLA pattern (Fig. 1e). In a similar manner, the π -bond order is also a measure of the double-bond character of a given carbon–carbon bond and the difference in the π -bond order between adjacent carbon–carbon bonds is the bond-order alternation (BOA). A BOA of about –0.55 is calculated for a donor–acceptor polyene with alternating double and single bonds⁹. Thus BOA and BLA are parameters that describe the mixing of the two resonance forms in the actual ground-state structure of the molecule^{7,8}.

To correlate the molecular β with the ground-state polarization and consequently with BLA, the molecule (CH₃)₂N–(CH=CH)₄–CHO was examined in the presence of an external static electric field, *F*, of varying strength, first using an AM1 hamiltonian⁷. Subsequently, it was examined at the semi-empirical ‘intermediate neglect of differential overlap–configuration interaction’ (INDO-CI) level, with the evaluation of the molecular polarizabilities through the ‘sum-over-states’ formulation at the correlated level⁸. In addition, similar studies have been performed using various other approaches and levels of theory^{10–13} which all arrived at essentially the same picture. In the INDO-CI study, the ground-state polarization was adjusted by applying an external static electric field of varying strength⁸. β was correlated with the average value of BOA. On going from the neutral polyene limit to the cyanine limit, β first increases, peaks in a positive sense for an intermediate structure, decreases and passes through zero at the cyanine limit (Fig. 2). From that limit and going to the charge-separated resonance structure, β continues to decrease and thus becomes negative, peaks in a negative sense, and then decreases again (in absolute value) to become smaller in the charge-separated structure. Evidence for the proposed relationships was derived from a study of a series of six molecules—that spanned a wide range of ground-state polarization and therefore BLA—that were examined in solvents of varying polarity¹⁴ by electric-field-induced second-harmonic generation (EFISH) experiments.

Optimizing β by design. The relative contribution of each limiting resonance structure to the ground-state structure of a molecule is related to their relative energies. When the two resonance forms are very different in energy the ground-state structure will be dominated by the lower-energy form and the molecules will exhibit a large degree of BLA (Fig. 1a, b). In contrast, if the two structures are

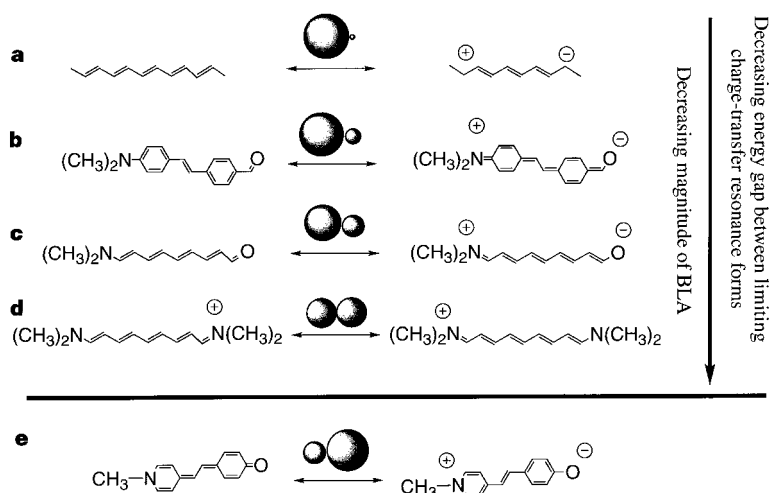


Figure 1 Illustration of the decrease in bond-length alternation (BLA; see text) as the two resonance forms contribute more equally to ground-state structure (a–d). The relative contribution of the resonance structures are schematically represented by the sizes of the balloons over the arrows. **b**, Molecule that neither gains nor loses aromatic stabilization on charge transfer; **c**, a molecule that loses aromatic stabilization on charge transfer and for which the sign of BLA is opposite that for the molecules in a–c.

the same in energy the molecule will exhibit very little BLA (Fig. 1d). Changing their relative contributions, and therefore the molecular structure, requires an understanding of the factors that influence the energies of the two forms. In organic molecules there are, to a first approximation, two factors that dominate the energetics of the resonance structures^{6,7}. First, there is a Coulombic term that is destabilizing when charge is separated (right-hand resonance structures shown in Fig. 1a, c). Increasing the strength of the donor and acceptor end groups, and/or placing the molecule in a more polar solvent can lead to stabilization of the charge-separated form. There is also an energy consideration associated with the topology of the molecule^{6,7}. If a molecule has six π -electrons in a ring, the molecule has an additional resonance stabilization and is referred to as aromatic. For molecules whose neutral forms are aromatic, charge separation will interrupt the aromaticity and yield structures with a higher-energy quinoidal resonance form. This disruption of the aromaticity results in additional destabilization due to the loss of aromatic stabilization⁶, and in such systems the molecule will be further biased towards the neutral resonance form (Fig. 1b)^{6,7}. If, on the other hand, the neutral form is quinoidal, then the molecule will gain aromatic stabilization in the charge-separated resonance form^{6,7}. This provides a strong driving force for charge separation. For example, the charge-separated form of the merocyanine (Fig. 1e) dominates the ground-state structure owing to the fact that it is aromatic. This illustrates that the key to controlling the ground-state structure is the use of a combination of donor-acceptor strength and bridge topology to alter the relative energetics of the two resonance forms.

Exploiting aromaticity. Until recently, most molecules that had been examined for nonlinear optics, such as donor-acceptor substituted stilbenes or diphenyl polyenes, had a very large BLA, typically greater than 0.10 Å. The calculations above, for example, predict that β is maximized at a value of about 0.04 Å; thus these molecules with large BLA are not sufficiently polarized to give the correct BLA needed to maximize β . It was proposed that the high value of BLA observed in the central polyene bridge of donor-acceptor substituted stilbenes and related molecules is indicative of an insufficient contribution of the charge-separated resonance forms to the ground-state configurations of the molecules, and is a consequence of the loss of aromatic stabilization in the charge-separated forms^{6,7}. To circumvent this problem, a strategy was proposed for the design of donor-acceptor polyenes in which the loss of aromaticity in one end (here the donor end) on charge separation would be somewhat offset by a gain in aromaticity in the opposite end on charge separation. It was predicted^{16–8} that this would lead to nearly the correct contribution of the charge-

separated form to the ground-state structure that is required to reach the BLA at which β is maximized. This hypothesis was confirmed using molecules containing acceptors whose topology dictates that aromaticity is gained on charge separation, such as 3-phenyl-5-isoxazolone (Fig. 3, compound 1) or *N,N'*-diethyl-thio-barbituric acid (Fig. 3, compound 2)¹⁵. These molecules have large values of β for their lengths. For second-order materials, which must be non-centrosymmetric as will be discussed in the following section, it is possible to induce polar order by poling the sample with an electric field; a large dipole moment, μ , in addition to β is desirable for this technique. It is therefore also useful to consider the product $\beta\mu$ as a molecular figure of merit. In the study of molecules containing 3-phenyl-5-isoxazolone or *N,N'*-diethyl-thio-barbituric acid, $\beta\mu$ values greater than $10,000 \times 10^{-48}$ e.s.u. (measured by EFISH at 1.9 μm) were reported for the first time. For comparison, $\beta\mu$ for DANS (a chromophore often used in poled polymer applications) is approximately 500×10^{-48} e.s.u. (at 1.9 μm)¹⁵.

An alternative approach to increasing the degree of ground-state polarization and thereby decreasing the magnitude of BLA is to replace strongly aromatic benzene rings with heteroaromatic rings that have smaller aromatic stabilization energies. This approach was successfully implemented^{16–18} using the low-aromaticity thiazole and thiophene rings (Fig. 3, compound 3). In particular, it has been shown^{17,18} that the replacement of both the benzene rings in DANS with thiophene rings results in a twofold increase in $\beta\mu$. This approach has been successfully applied to many systems; one important feature of this approach is that the heterocyclic analogues of stilbene compounds not only give rise to large nonlinearities, but also in many cases have adequate thermal stability for use in polymers which may have processing conditions well above 200 °C. Thermal stability of chromophores was addressed in another manner. It was recently shown¹⁹ that for many chromophores with amino donors, replacement of the alkyl functionality on the nitrogen with an aryl group (like phenyl) could improve the thermal stability dramatically (Fig. 3, compound 4). Moreover, it is possible to combine the approaches to yield chromophores with large β and excellent thermal stability²⁰.

Several workers have recently examined compounds in which the molecule gains aromatic stabilization on charge separation. Such molecules exhibit negative β because their excited-state dipole moments are lower than the ground-state dipole moment. As these molecules have a large degree of charge separation in the ground state, they tend to have large dipole moments relative to stilbene-like molecules. A particularly striking example is compound 5 (Fig. 3), for which $\beta\mu$ is $-9,500 \times 10^{-48}$ e.s.u. (at 1.9 μm)

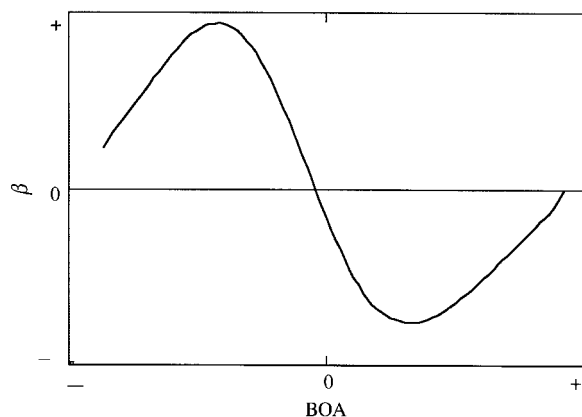


Figure 2 Plot of β (see text) versus bond-order alternation (BOA; see text) for a simple donor-acceptor polyene, $(\text{CH}_3)_2\text{N}-(\text{CH}=\text{CH})_4-\text{CHO}$, polarized to increasing extents by application of an external field along the dipole vector of the molecule.

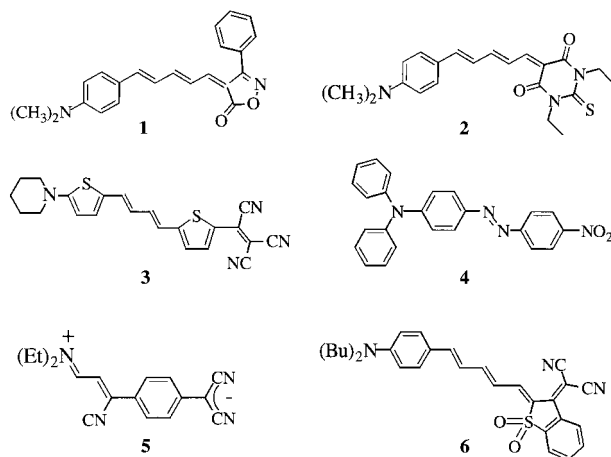


Figure 3 Structures and numbering scheme for several highly nonlinear second-order NLO chromophores (see text).

and μ is 45 D²¹. This is an astonishingly large dipole moment compared with DANS for which μ is¹⁵ roughly 7 D.

Thus it has been shown that by careful manipulation of both donor–acceptor strength and bridge topology it is possible to produce a range of compounds, from neutral polyene-like molecules through cyanine-like molecules to highly charge-separated molecules. Given the flexibility that chemists have in changing the extent of ground-state polarization and BLA, it is possible to ‘tune’ to the point on the BLA or polarization coordinate where β (or $\beta\mu$) is optimized for a given length of molecule.

From molecules to bulk materials

As noted earlier, second-order NLO effects occur only in molecules lacking a centre of symmetry. Likewise, a bulk material comprised of nonlinear molecules must also lack a centre of symmetry if the molecular β is to lead to a macroscopic nonlinearity^{1,2}. For electro-optic applications, and for some photorefractive applications, the polymers must also show orientational stability. Thus, once the chromophores are orientated in the polymer such that the material is non-centrosymmetric (as will be described below), it is necessary that the orientational distribution of the chromophores cannot be allowed to randomize over time, even at elevated temperatures, or the second-order nonlinearity will decay. In contrast to electro-optic applications, many photorefractive applications do not require long-term orientational stability. Accordingly, we will describe some of the advances made for: (1) achieving stable orientational order for electro-optic polymers; and (2) optimizing the efficiency of photorefractive materials where orientational stability is not an issue. Many materials issues are common to both applications. For example, the materials must have low optical losses from either absorption or scattering, and they must be environmentally and photochemically stable if they are to be of

practical use. Furthermore, if the materials are to be incorporated into devices it is necessary that they can be processed. Thus the task of designing useful materials requires an understanding of the interplay and trade-offs among these various parameters which are ultimately defined by the specific application.

Electro-optic polymers. The method most widely used to impart non-centrosymmetry in non-crystalline systems is the poled-polymer approach^{22,23}. If dipolar NLO species are dissolved in a polymer and subjected to a large electric field at or above the temperature at which the polymer becomes rubbery (called its glass transition temperature, T_g), the interaction of the dipole with the field causes the dipolar species to orientate, to a certain extent, in the direction of the applied field. If the polymer is cooled back to the glassy state with the field applied, then the field-induced non-centrosymmetric orientation can be frozen in place, yielding a material with a second-order optical nonlinearity²⁴. However, this thermodynamically unstable orientation quickly decays in low- T_g polymers such as poly(methylmethacrylate), resulting in a greatly reduced nonlinearity. When an electric field is applied to such a poled polymer, the index of refraction of the material will change. This sensitivity of the material to having its index of refraction changed by application of an external electric field is characterized by its so-called electro-optic coefficient (r_{33}). For many years it had been suggested that electro-optic polymers could show values of r_{33} much bigger than that of a technologically important crystal, lithium niobate. Recently, using a highly nonlinear molecule (compound **6** in Fig. 3) in polycarbonate (an intermediate- T_g polymer), it was shown²⁵ that a very large r_{33} of 55 pm V⁻¹ measured at 1.3 μ m was achievable. This is nearly twice the value for lithium niobate, 30.8 pm V⁻¹. But the **6**-containing polymer had only moderate thermal stability, and so here we focus on more recent designs that more effectively lock the orientated molecules in place, even

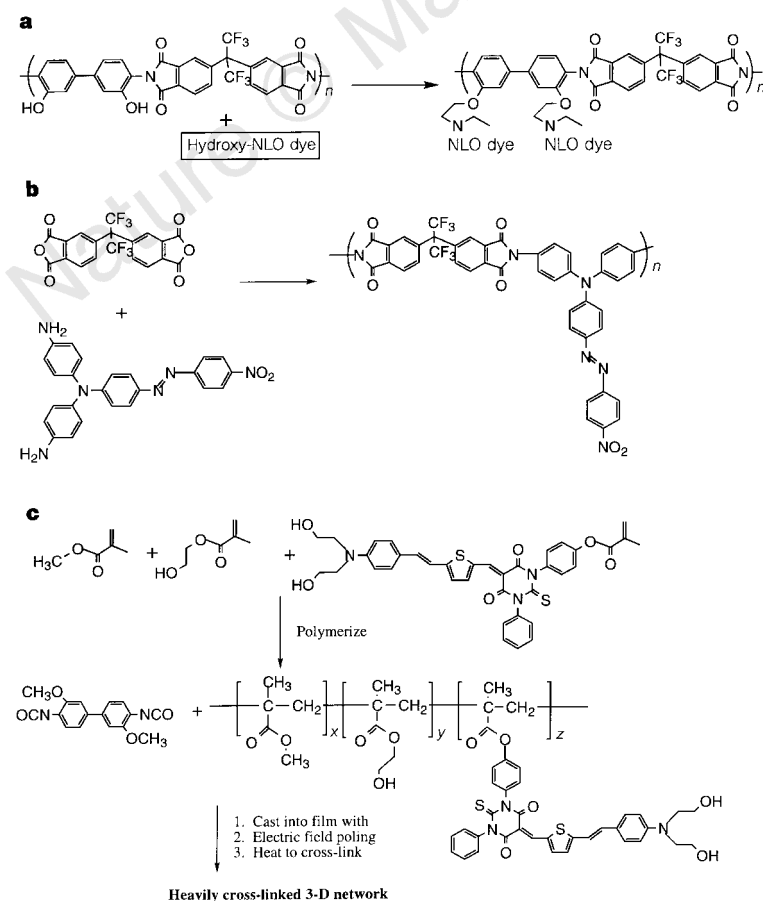


Figure 4 a, High-temperature NLO side-chain aromatic polyimides synthesized by using a post-Mitsunobu reaction to graft NLO chromophores onto hydroxyl-containing precursor polyimides. **b**, Exceptionally thermally stable NLO side-chain aromatic polyimides made by using a 'donor-embedded' approach containing no flexible connectors or tethers to the NLO chromophores. **c**, Heavily cross-linked three-dimensional NLO polyacrylates synthesized by 'double-ended' crosslinking of the hydroxyl-containing precursor polyacrylates during high-electric-field poling process.

at elevated temperatures, and that incorporate highly nonlinear molecules.

Researchers have adopted two approaches to improve the long-term stability of molecular orientation in poled polymers. First, several groups have used very high- T_g polymers as host matrices. The rationale is that if the T_g of the polymer is roughly 150 to 200 °C above the ultimate operating temperature, decay of the orientation would be negligible over the device lifetime. This was first demonstrated with a commercial polyimide and NLO chromophore²⁶, and more recently with highly nonlinear heteroaromatic chromophores^{27–29}. A.K.-Y.J. and co-workers have developed classes of chromophores based on thiophene conjugated units substituted with donor and acceptor groups. The chromophore was stable to both imidization (30 min at 200 °C) and poling (10 min at 220 °C)²⁷. The polymer showed long-term stability at elevated temperatures of 120 °C (150 °C) for more than 30 days, with the electro-optic coefficient maintained at 80% (60%) of its initial value. Similar molecules were incorporated as guests in polyquinoline thin films²⁹. After poling, the electro-optic coefficient remained stable at 26 pm V⁻¹ at 80 °C for more than 80 days.

The motion of chromophores in polymer matrices might be further restricted if the chromophore were covalently attached at one or more sites to the polymer (Fig. 4a). Accordingly, several groups have explored chromophores attached covalently to polyimides either in the main chain or as a side chain. This

approach can lead to polymers with exceptional thermal stability. A.K.-Y.J. and co-workers have recently reported a two-step synthesis of NLO side-chain aromatic polyimides³⁰. The advantages of this procedure include the ease of controlling the loading level of the chromophore and of adjusting the polymer backbone rigidity³⁰.

It seems that the ultimate achievable stability is related to the inherent flexibility of the polymer backbone and the degree of freedom in the linkage between the chromophore and the polymer (that is, whether the chromophore is attached in the main chain or in a side chain). Thus systems with rigid backbones and short rigid linkers will tend to have the highest T_g and also the best temporal stability. Also, if the rigidity leads to improved temporal stability, it may lead to some degree of crystallinity which can result in significant optical losses due to scattering. Recently the preparation was reported of a soluble NLO substituted polyimide derivative containing this type of diaryl amino donor head embedded in a polyimide main chain^{31,32}. The resulting polymer is characterized by an extraordinary stability of the induced polar order, due partially to a high T_g (~350 °C) of the system (Fig. 4b). But it is not surprising that in this polymer, even before poling, optical losses due to scattering were relatively large.

As might be expected, the continuous increases of the rigidity and T_g of poled polymers imposes constraints on the selection of suitable chromophores that can survive the high-temperature poling and processing conditions. To circumvent this problem, an alternative approach would be to develop a lower- T_g material which contains reactive functional groups on the chromophore or the polymer backbone, can be processed into optical-quality thin films, and can then be poled and crosslinked simultaneously at higher temperatures to 'lock' the chromophores and polymer in place to achieve better control over the poling-induced polar order. Care must be taken to ensure that the densification of the lattice is not effected before a reasonable degree of polar order is introduced by electric field poling. Initially, the crosslinking approach for making electro-optic materials was exploited by Marks and co-workers³³ and an IBM group³⁴ using chromophore-containing poly(*p*-hydroxystyrene) and chromophore-embedded epoxy as thermosetting matrices. The resulting crosslinked films are impervious to organic solvents and possess good orientational stability at ambient temperature. In addition to the epoxy resins, a number of new approaches including the modified polyurethanes^{35,36} and polyimides^{37,38} have been used to further enhance orientational stability of the poled polymers at temperatures ranging from 85 to 150 °C. Dalton *et al.*³⁹ have recently made important advances in poled materials by use of a 'double-ended' crosslinking approach in which the chromophores containing reactive functionalities are at both donor and acceptor ends. Sufficiently different reactive functionalities allow a two-step process selectively to polymerize the precursor polymer first, and then perform the hardening step during the curing/poling process (Fig. 4c). The group has reported³⁹ long-term stability at 90 °C for more than 40 days, with r_{33} values ranging from 9 to 17 pm V⁻¹.

Photorefractive polymers. During the past two decades, the design of NLO chromophores has been mainly aimed at optimizing the performance of electro-optic polymers. With the recent development of highly efficient photorefractive polymers, interest in chromophore design has gained additional momentum and new design strategies have emerged. Whereas the first photorefractive polymers⁴⁰ combined solely electro-optic and photoconducting properties (as in their inorganic crystalline counterparts such as LiNbO₃), present-day polymers show additional orientational effects⁴¹. These orientational effects are largely responsible for the high performance of this new class of materials, and have drastically changed the optimization criteria of chromophores for incorporation into low- T_g photorefractive polymers. Here again, the BLA/BOA theory seems to provide a powerful tool

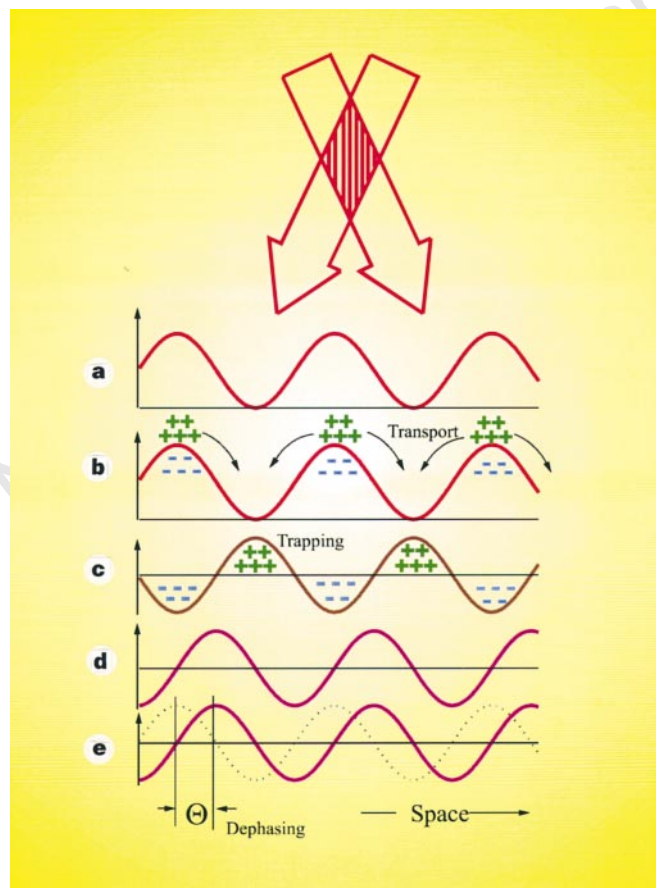


Figure 5 Illustration of the photorefractive effect. The overlap of two coherent laser beams creates a spatially varying optical interference pattern (a). In the high-intensity regions, charge carriers are generated (b). One type of carrier is transported and trapped (c), creating an alternating space-charge field (d). The space-charge field induces a refractive-index grating via the electro-optic effect (e). This grating (solid curve) is phase-shifted (by Θ) with respect to the initial light distribution (dotted curve).

for optimizing chromophores for photorefractive applications. To gain a better understanding of these new design criteria, we review briefly some of the basic principles of photorefractive polymers.

The photorefractive effect is based on a combination of photoconducting and electro-optic properties, and can lead to high refractive index variations under the illumination of low-power lasers (Fig. 5). When two laser beams intersect in such a material, they create an interference pattern with a strongly nonuniform distribution of light intensity. Excess charges generated through optical absorption in the high-intensity regions migrate to the regions of low intensity, leading to charge separation and the build-up of an internal electric field. Because the material is electro-optic, the internal electric field changes internally the refractive index of the material. As a result, the initial light distribution is optically encrypted in the form of a refractive-index pattern in the material. Owing to trapping of the carriers, the internal space-charge field and the corresponding refractive-index grating have a lifetime ranging from nanoseconds to several years depending on the photorefractive material. Thus photorefractive materials are suitable for the recording and storage of optically encoded information such as holograms. In addition, photorefractive materials are reconfigurable: recording of optical information can be performed in real time. The redistribution of photogenerated charges follows any changes occurring in the optical wavefront during recording. Hence photorefractive materials are not only suitable for optical storage but also show particular promise for real-time optical processing applications⁴². During the past thirty years, the photorefractive effect has been studied mainly in inorganic crystals that are difficult to produce and to process. In these traditional photorefractive materials, the refractive-index modulation arises only from the space-charge field acting on the electro-optic (Pockels) effect of the materials. Thus the molecular figure of merit developed for early photorefractive polymers, such as the first photorefractive polymer developed at IBM⁴⁰, was related to $\beta\mu/M_r$, as discussed above for purely electro-optic polymers (where M_r is the relative molecular mass). In contrast,

highly efficient photorefractive polymers, such as the plasticized polyvinyl carbazole based polymers incorporating 2,5-dimethyl-4-(*p*-phenyl-azo)anisole (DMNPAA) developed at the University of Arizona⁴¹, have a low T_g and exhibit an orientational contribution to the refractive-index modulation. The so-called orientational enhancement effect⁴³ is due to the ability of the chromophores to orientate at room temperature under the influence of the total electric field which is the superposition of the internal modulated field and the externally applied field. As a result, in steady state after photorefractive hologram formation, the molecules no longer have a uniform orientation but can have an orientation that is spatially modulated both in magnitude and in direction. That periodic orientation of the chromophores doubles the effect of the electro-optic contribution and, more importantly, leads to a modulated birefringence that significantly enhances the total refractive-index modulation⁴³. This effect occurs, however, only in polymers with a T_g below, or close to, the operating temperature. For such photorefractive polymers, a possible molecular figure of merit is $A(T)\Delta\alpha\mu^2 + \beta\mu$ (where $\Delta\alpha$ is the difference between the molecule's polarizability along its molecular axis versus smaller off-axis contributions, and $A(T)$ is the scaling factor $2/9kT$ where k is Boltzman's constant and the temperature T is 300 K)⁴⁴. Thus in designing molecules, not only do the values of the dipole moment and the hyperpolarizability play a role, but so too does the polarizability anisotropy, $\Delta\alpha$, of the molecule.

The ability of the BLA/BOA theory to help in tailoring these three molecular constants is extremely important for the design of optimized photorefractive polymers. We have analysed calculations performed on the molecule $(CH_3)_2N-(CH=CH)_4CHO$ in the presence of an external field, where the field is used to modulate the ground-state polarization and concomitantly BOA as described above; we have plotted $A(T)\Delta\alpha\mu^2$ and $\beta\mu$ as a function of BOA (Fig. 6)⁴⁴. In the region where $A(T)\Delta\alpha\mu^2$ and $\beta\mu$ are peaked, the magnitude of $A(T)\Delta\alpha\mu^2$ is over 80 times higher than that of $\beta\mu$. As a result, the contribution of the $\beta\mu$ term is almost negligible and $A(T)\Delta\alpha\mu^2 + \beta\mu \approx A(T)\Delta\alpha\mu^2$ for this molecule. Thus these calculations showed⁴⁴ that linear polymethine molecules should be opti-

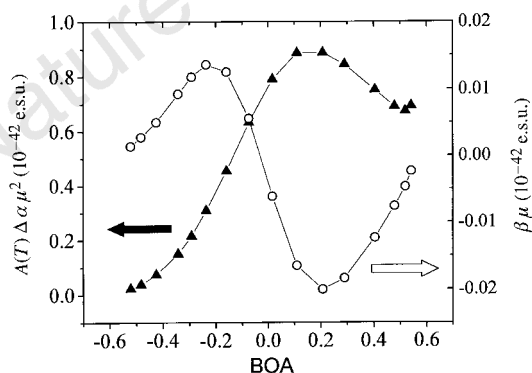


Figure 6 Plot of $A(T)\Delta\alpha\mu^2$ and $\beta\mu$ (see text) versus bond-order alternation (BOA) for $(CH_3)_2N-(CH=CH)_4-CHO$.

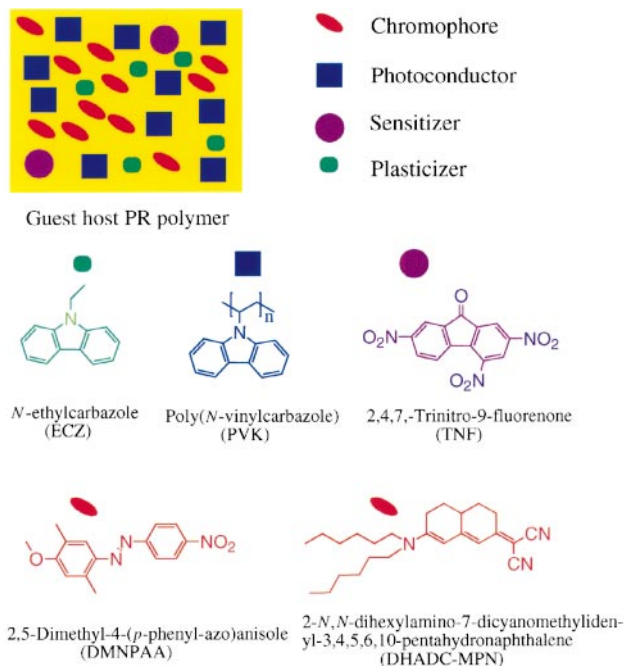


Figure 7 Comparison of low- T_g photorefractive (PR) polymers, illustrating the various components that impart photoconductivity, plasticization, sensitization and orientational nonlinearity. T_g is the glass transition temperature.

mized for photorefractive applications when they are polarized somewhat beyond the cyanine limit⁴⁴. Based on these results, we synthesized the polyene molecule 2-*N,N*-dihexylamino-7-dicyanomethylidenyl-3,4,5,6,10-pentahydronaphthalene (DHADC-MPN) (Fig. 7). The molecule was used as a dopant molecule in mixtures of polyvinyl carbazole and ethylcarbazole that act as the photoconducting matrix (Fig. 7). The photorefractive properties were tested by four-wave mixing and two-beam coupling experiments. In four-wave mixing experiments, maximum diffraction was observed in 105- μm -thick samples at an applied field of $30\text{ V }\mu\text{m}^{-1}$ at 633 nm, corresponding to an improvement of Δn by a factor of four compared to previous DMNPAA-based polymers (Fig. 7) with 50 wt% chromophore loading. In addition, the DHADC-MPN-based samples remain homogeneous and thus retain their excellent optical properties over long periods of time.

Overall, the most efficient photorefractive polymers to date are plasticized guest/host systems. However, as for purely electro-optic polymers, there is a strong need for photorefractive polymers with a high T_g . Such polymers are poled and have stable electro-optic properties, but they do not show the orientational enhancement effect discussed above. These materials will be of use for long-term holographic data storage. Thus their optimization is similar to that of purely electro-optic polymers; recent progress in developing chromophores with $\beta\mu > 10,000 \times 10^{-48}$ e.s.u. (at 1.9 μm) will probably provide new opportunities for the fabrication of efficient, high- T_g photorefractive polymers.

Into the marketplace

The field of organic NLO has advanced rapidly in the past five years. Large steps have been made in understanding the origins of the NLO response, and optimizing both molecular and polymeric electro-optic and photorefractive responses. Opto-electronic devices based on polymeric materials are starting to become commercially available. The low cost and ease of processing offered by polymers make them extremely competitive with inorganic systems. Examples of organic devices already in the market place are the thermo-optic switches (which face several of the same technological issues as the electro-optic polymers) offered by Akzo-Nobel. The materials requirements for electro-optic applications are rather more severe. However, questions of stability have been addressed, and it seems that systems with an adequate combination of transparency, thermal stability and optical nonlinearities now exist for certain applications. Although problems related to scale-up of materials synthesis, device fabrication and system integration still exist, TACAN Corporation (Carlsbad, CA) has recently announced that they are incorporating a polymeric electro-optic modulator into a commercial transmission system to test field applications. In the area of organic photorefractive polymers, useful devices have been demonstrated⁴², and advances in materials show no signs of reaching a plateau. □

S. R. Marder is at the Beckman Institute, and the Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California 91125, USA. B. Kippelen and N. Peyghambarian are at the Optical Sciences Center, University of Arizona, Tucson, Arizona 85721, USA. A. K.-Y. Jen is at the Department of Chemistry, Northeastern University, Boston, Massachusetts 02115, USA

- Chemla, D. S. & Zyss, J. (eds) *Nonlinear Optical Properties of Organic Molecules and Crystals* (Academic, Orlando, 1987).
- Marder, S. R., Sohn, J. E. & Stucky, G. D. (eds) *Materials for Nonlinear Optics: Chemical Perspectives* (Vol. 455, ACS Symp. Ser., Am. Chem. Soc., Washington DC, 1991).
- Peyghambarian, N., Koch, S. W. & Mysyrowicz, A. *Introduction to Semiconductor Optics* (Prentice-Hall, Englewood Cliffs, 1993).
- Buckingham, A. D. Permanent and induced molecular moments and long-range intermolecular forces. *Adv. Chem. Phys.* **12**, 107–142 (1967).
- Oudar, J. L. & Chemla, D. S. Hyperpolarizabilities of the nitroanilines and their relations to the excited state dipole moment. *J. Chem. Phys.* **66**, 2664–2668 (1977).
- Marder, S. R., Beratan, D. N. & Cheng, L. T. Approaches for optimizing the 1st electronic hyperpolarizability of conjugated organic molecules. *Science* **252**, 103–106 (1991).
- Gorman, C. B. & Marder, S. R. An investigation of the interrelationships between linear and nonlinear polarizabilities and bond-length alternation in conjugated organic molecules. *Proc. Natl Acad. Sci. USA* **90**, 11297–11301 (1993).

- Meyers, F., Marder, S. R., Pierce, B. M. & Brédas, J. L. Electric-field modulated nonlinear-optical properties of donor-acceptor polyenes: sum-over-states investigation of the relationship between molecular polarizabilities (alpha, beta, and gamma) and bond-length alternation. *J. Am. Chem. Soc.* **116**, 10703–10714 (1994).
- Salem, L. *The Molecular Orbital Theory of Conjugated Systems* (Benjamin, New York, 1966).
- Kuhn, C. Step potential model for nonlinear optical-properties of polyenes, push-pull polyenes and cyanines and the motion of solitons in long-chain cyanines. *Synth. Met.* **41–43**, 3681–3688 (1991).
- Chen, G. H. & Mukamel, S. Nonlinear susceptibilities of donor-acceptor conjugated systems-coupled-oscillator representation. *J. Am. Chem. Soc.* **117**, 4945–4964 (1995).
- Dehu, C. *et al.* Solvent effects on the second-order nonlinear-optical response of pi-conjugated molecules—a combined evaluation through self-consistent reaction field calculations and hyper-Rayleigh scattering measurements. *J. Am. Chem. Soc.* **117**, 10127–10128 (1995).
- Lu, D. Q., Chen, G. H., Perry, J. W. & Goddard, W. A. Valence-bond charge-transfer model for nonlinear-optical properties of charge-transfer organic-molecules. *J. Am. Chem. Soc.* **116**, 10679–10685 (1994).
- Bourhill, G. *et al.* Experimental demonstration of the dependence of the 1st hyperpolarizability of donor-acceptor-substituted polyenes on the ground-state polarization and bond-length alternation. *J. Am. Chem. Soc.* **116**, 2619–2620 (1994).
- Marder, S. R. *et al.* Large 1st hyperpolarizabilities in push-pull polyenes by tuning of the bond-length alternation and aromaticity. *Science* **263**, 511–514 (1994).
- Dirk, C. W., Katz, H. E., Schilling, M. L. & King, L. A. Use of thiazole rings to enhance molecular second-order nonlinear optical susceptibilities. *Chem. Mater.* **2**, 700–705 (1990).
- Jen, A. K.-Y., Rao, V. P., Wong, K. Y. & Drost, K. J. Functionalized thiophenes—second-order nonlinear optical-materials. *J. Chem. Soc. Chem. Commun.* 90–92 (1993).
- Rao, V. P., Jen, A. K. Y., Wong, K. Y. & Drost, K. J. Dramatically enhanced second order nonlinear-optical susceptibilities in tricyanovinylthiophene derivatives. *J. Chem. Soc. Chem. Commun.* 1118–1120 (1993).
- Moylan, C. R. *et al.* Nonlinear-optical chromophores with large hyperpolarizabilities and enhanced thermal stabilities. *J. Am. Chem. Soc.* **115**, 12599–12600 (1993).
- Jen, A. K.-Y., Cai, Y. M., Bedworth, P. V. & Marder, S. R. Synthesis and characterization of highly efficient and thermally stable diphenylamino-substituted thiophene stilbene chromophores for nonlinear-optical applications. *Adv. Mater.* **9**, 132–135 (1997).
- Cross, G. H., Bloor, D. & Szablewski, M. "Blue Window" organics for frequency doubling devices. *Nonlin. Optics* **14**, 219–223 (1995).
- Burland, D. M., Miller, R. D. & Walsh, C. A. Second-order nonlinearity in poled-polymer systems. *Chem. Rev.* **94**, 31–75 (1994).
- Dalton, L. R. *et al.* Polymeric electro-optic modulators: materials synthesis and processing. *Adv. Mater.* **7**, 519–540 (1995).
- Singer, K. D., Sohn, J. E. & Lalama, S. J. Second harmonic generation in poled polymer films. *Appl. Phys. Lett.* **49**, 248–250 (1986).
- Ahlheim, M. *et al.* Chromophores with strong heterocyclic acceptors: a poled polymer with a large electro-optic coefficient. *Science* **271**, 335–337 (1996).
- Wu, J. W. *et al.* Thermal stability of electro-optic response in poled polyimide systems. *Appl. Phys. Lett.* **58**, 225–227 (1991).
- Jen, A. K.-Y., Wong, K. Y., Rao, V. P., Drost, K. & Cai, Y. M. Thermally stable poled polymers: highly efficient heteroaromatic chromophores in high temperature polyimides. *J. Electron. Mater.* **23**, 653–657 (1994).
- Wong, K. Y. & Jen, A. K. Y. Thermally stable poled polyimides using heteroaromatic chromophores. *J. Appl. Phys.* **75**, 3308–3310 (1994).
- Cai, Y. M. & Jen, A. K. Y. Thermally stable polyquinoline thin film with very large electro-optic response. *Appl. Phys. Lett.* **67**, 299–301 (1995).
- Chen, T. A., Jen, A. K.-Y. & Cai, Y. M. Facile approach to nonlinear optical side-chain aromatic polyimides with large second-order nonlinearity and thermal stability. *J. Am. Chem. Soc.* **117**, 7295–7296 (1995).
- Verbiest, T. *et al.* Exceptionally thermally stable polyimides for second-order nonlinear optical applications. *Science* **268**, 1604–1606 (1995).
- Miller, R. D. *et al.* Donor-embedded nonlinear optical side chain polyimides containing no-flexible tether: materials of exceptional thermal stability of electrooptic applications. *Macromolecules* **28**, 4970–4974 (1995).
- Hubbard, M. A., Marks, T. J., Yang, J. & Wong, G. K. Poled polymeric nonlinear optical materials. Enhanced second harmonic generation stability of cross-linkable matrix/chromophore ensembles. *Chem. Mater.* **1**, 167–169 (1989).
- Eich, M., Björklund, G. C. & Yoon, D. Y. Poled amorphous polymers for second-order nonlinear optics. *Polym. Adv. Technol.* **1**, 189–198 (1990).
- Chen, M., Yu, L., Dalton, L. R., Shi, Y. & Steier, W. H. New polymers with large and stable second-order nonlinear optical effects. *Macromolecules* **24**, 5421–5428 (1991).
- Boogers, J. A. E., Klasse, P. Th. A., de Wlieger, J. J. & Tinemans, A. H. A. Cross-linked polymer materials for nonlinear optics 2: polyurethanes bearing azobenzene dyes. *Macromolecules* **27**, 205–209 (1994).
- Lin, J. T., Hubbard, M. A., Marks, T. J., Lin, W. & Wong, G. K. Poled polymeric nonlinear optical materials. Exceptional second harmonic generation temporal stability of a chromophore-function-alized polyimide. *Chem. Mater.* **4**, 1148–1150 (1992).
- Topolsky, G., Lecomte, J. P. & Meyrueix, R. Maleimide-based cross-linkable electrooptic polymers with excellent thermal stability characteristics. *Macromolecules* **26**, 7383–7385 (1993).
- Yu, L. *et al.* Thermally curable second-order nonlinear-optical polymers. *Appl. Phys. Lett.* **60**, 1655–1657 (1992).
- Ducharme, S., Scott, J. C., Twieg, R. J. & Moerner, W. E. Observation of the photorefractive effect in a polymer. *Phys. Rev. Lett.* **66**, 1846–1849 (1991).
- Meerholz, K., Volodin, B. L., Sandalphon, Kippelen, B. & Peyghambarian, N. A. Photorefractive polymer with high optical gain and diffraction efficiency near 100%. *Nature* **371**, 497–500 (1994).
- Volodin, B. L., Kippelen, B., Meerholz, K., Javid, B. & Peyghambarian, N. A. Polymeric optical pattern-recognition system for security verification. *Nature* **383**, 58–60 (1996).
- Moerner, W. E., Silence, S. M., Hache, F. & Björklund, G. C. Orientationally enhanced photorefractive effect in polymers. *J. Opt. Soc. Am. B* **11**, 320–330 (1994).
- Kippelen, B., Meyers, F., Peyghambarian, N. & Marder, S. R. Chromophore design for photorefractive applications. *J. Am. Chem. Soc.* **119**, 4559–4560 (1997).

Acknowledgements. We thank the many authors cited in this Review, as well as those whose contributions could not be discussed owing to length constraints, for their contributions to the field and to our understanding of it; and J. Perry and F. Meyers for help in the preparation of this manuscript. This work was supported by the US Office of Naval Research, the US Air Force Office of Scientific Research, the National Science Foundation, and the US Ballistic Missiles Defense Organization.

A direct image of the obscuring disk surrounding an active galactic nucleus

Jack F. Gallimore*, Stefi A. Baum† & Christopher P. O'Dea†

* Max-Planck-Institut für extraterrestrische Physik, Postfach 1603, D-85740 Garching bei München, Germany

† Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, Maryland 21218, USA

Active galactic nuclei (AGN) are generally accepted to be powered by the release of gravitational energy in a compact accretion disk surrounding a massive black hole^{1,2}; such disks are also thought necessary to collimate the powerful radio jets seen in some AGN³. The unifying classification schemes for AGN further propose that differences in their appearance can be attributed to the opacity of the accreting material, which may obstruct our view of the central region of some systems. The popular model for the obscuring medium is a parsec-scale disk of dense molecular gas⁴, although evidence for such disks has been mostly indirect, as their angular size is much smaller than the resolution of conventional telescopes. Here we report direct images of a parsec-scale disk of ionized gas within the nucleus of NGC1068, the archetype of obscured AGN. The disk is viewed nearly edge-on, and the ionized clouds observed within the ionized disk are opaque to high-energy radiation, consistent with the unifying classification schemes. The projected axes of the disk and AGN are aligned, from which we infer that the ionized gas disk traces the outer regions of the long-sought inner accretion disk.

The nucleus of the galaxy NGC1068 hosts the archetypal example of an obscured AGN. The popular model for the obscuring medium is a parsec-scale, molecular disk surrounding the AGN⁵, perhaps ultimately feeding an accretion disk^{6,7}. One difficulty for observational tests has been that the location of the obscured, central ionizing source is unknown. It has been argued on several grounds that the radio source S1 marks the location of the hidden AGN in NGC1068 (refs 8, 9). Located at the southern end of the arcsecond-scale radio jet, S1 is an unusual radio source in two respects. First, in contrast with the rest of the radio jet, its radio spectrum is relatively flat (spectral index $\alpha = +0.3$; $S_\nu \propto \nu^\alpha$ (ref. 8)), and second, it is a source of H₂O and OH maser emission¹⁰, which distinguishes regions of peculiarly warm (1,000 K) and dense ($\geq 10^8$ molecules cm⁻³) molecular gas. We argued that S1 might trace emission from molecular clouds defining the inner surface of the proposed obscuring disk, whose surfaces would be exposed directly to the central X-ray source and are therefore hot and highly ionized⁸.

We made two predictions for very-long-baseline interferometry (VLBI) observations⁸. First, S1 should resolve into a parsec-scale, linear radio structure, tracing the profile of an edge-on disk or 'torus' projected onto the sky, and located within the warm, molecular disk mapped in part by H₂O masers¹⁰. Second, the mean surface brightness of S1, in temperature units corresponding to an equivalent blackbody radiator (brightness temperature), should be $T_b \approx 10^6$ K for scattering-diffused emission or thermal free-free emission⁸.

To test these predictions, we have imaged the subarcsecond radiostructure of NGC1068 using the 10-station Very Large Baseline Array (VLBA), augmented by the phased, 26-element Very Large Array (VLA). The new images are displayed in Fig. 1. A single, deep (8.8 h on-source) integration was obtained at 8.4 GHz only. The observations and data reduction followed standard techniques with

exceptions as follows. On continental-scale baselines, NGC1068 is not sufficiently bright at 8 GHz to calibrate interferometric fringes within averaging times comparable to the atmospheric coherence time. Instead, we measured fringe-rate and fringe-delay corrections from short scans of the nearby calibrator source 0237 – 027. Less than 0.2% of the data (one out of a total of nearly 600 baseline-hours) were affected by phase rotations resulting from fringe solution ambiguities, and the radio sources S1, C and NE were

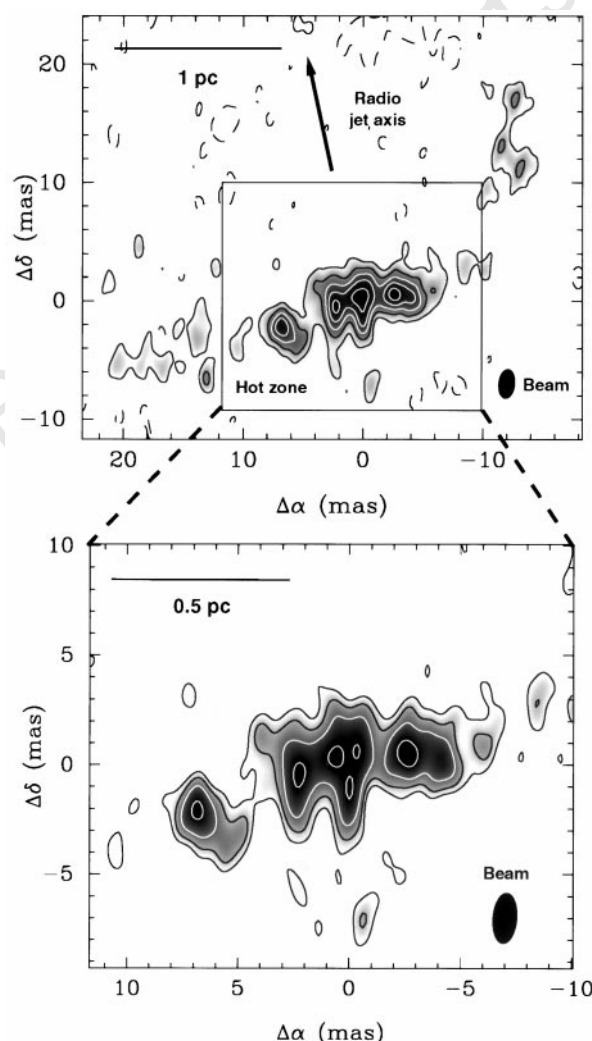


Figure 1 VLBI images of the radio component S1 of NGC1068. The total recovered flux of the 'hot zone' (HZ) is 6.9 mJy, or ~60% of the flux anticipated by a power-law interpolation of the 5-GHz and 22-GHz measurements of Gallimore *et al.*¹⁴. In contrast, less than a total of ~1 mJy arises from compact structures lying outside the HZ but within S1. Owing to an instability in the deconvolution algorithm used to produce this image, some of the compact sources may be artificially enhanced at the expense of the diffuse emission²⁸. The compact sources are nevertheless real, as they are also distinguishable on the unprocessed images. δ , declination; α , right ascension. Top panel, naturally weighted image of S1; the beam size (full-width at half-maximum, FWHM), indicated by the black ellipse in the lower right-hand corner, is 2.5×1.4 mas. We have marked and labelled the HZ and the local jet axis towards radio jet component C. Note that, in projection, the extent of the HZ and the direction of the radio jet are at right angles to each other, suggesting a common symmetry axis. Scaled logarithmically, the contour levels are ± 0.10 (2.5 σ), 0.22, 0.35, 0.47 and 0.59 mJy beam⁻¹, or $\pm 0.49, 1.1, 1.7, 2.3$ and 2.9 in brightness temperature units of 10^6 K. Bottom panel, uniformly weighted image of the HZ; the beam size (FWHM) is 2.3×1.1 mas. The contour levels are ± 0.16 (2.5 σ), 0.24, 0.36 and 0.54 mJy beam⁻¹, or $\pm 1.1, 1.6, 2.5, 3.7$ in units of 10^6 K.

clearly detected on the initial maps. The resulting, fringe-calibrated data were coherent over sufficiently long intervals to allow self-calibration, and so we removed the residual phase wraps using five iterations of phase-only self-calibration. In order to focus specifically on the nuclear emission, the VLBA images of NE and C will be presented elsewhere. Here we focus on the 'hot zone' (HZ), the brighter, central region of S1, and also defer discussion of fainter radio emission to future work.

The HZ comprises nine distinguishable compact sources, each of total flux density $S_\nu \approx 0.65$ mJy ($1 \text{ mJy} = 10^{-26} \text{ erg s}^{-1} \text{ cm}^{-2} \text{ Hz}^{-1}$), embedded in diffuse emission. These observations only marginally resolve the individual compact sources. Based on gaussian model fits and image moment analysis, the deconvolved source sizes are typically ~ 1 milliarcsecond (mas), or ~ 0.07 pc at the distance of NGC1068. These measurements are, however, uncertain owing to confusion between neighbouring sources, blurring due to residual phase errors, and possible enhancement by deconvolution. We also estimated limits on the source sizes based on inspection of the interference fringes. Less than half (~ 3 mJy) of the recovered flux of the HZ is detected on baselines corresponding to angular sizes $< 2 \times 1$ mas; at least half of the flux from the HZ must therefore arise from structures ≥ 1 mas in size, consistent with measurements of the synthesized image. This limit is conservative, as 1–2 mJy worth of milliarcsecond-scale components are also detected towards components NE and C. Any one or two of these compact sources may be smaller than 1 mas, but this caveat will not affect the main conclusions.

The compact sources trace a slightly curving line along a position angle of 110° , measured east of north. The HZ is therefore nearly at right angles to the collimation axes defined by the local radio jet and polarization axes, the latter describing the collimation axis for escaping ionizing radiation¹¹. This geometry is fully consistent with our prediction that the radio emission from S1 traces emission not from a streaming jet but rather from gas in an ionized disk surrounding the AGN. We next consider the implications of this disk model, making the simplifying assumption that the HZ gas lies at a common radius, ~ 0.3 – 0.5 pc, from the AGN.

Opaque synchrotron emission is the conventional explanation for flat-spectrum radio sources, but the brightness temperatures over the HZ are too low for synchrotron self-absorption¹². There are two likely alternatives, illustrated in Fig. 2, each being a variation on emission from ionized gas in a disk⁸. The first is synchrotron emission from a synchrotron-opaque, compact radio source, presumably the AGN, which is not viewed directly but in reflection by electron-scattering from the ionized gas disk.

The limits for this model are set by requiring that the electron scattering opacity (τ_e) must exceed the opacity to free-free absorption (τ_{ff}), and that the hidden radio source must not be so luminous that it would have been detected in reflection on larger scales. Based on the sensitivities of our radio continuum images⁸ and the reflecting properties of the electron-scattering mirror¹³, we estimate that the hidden radio source can be no brighter than $S_\bullet \approx 3.5$ Jy. We estimated limits on the plasma properties by exploring a grid of electron densities n_e , electron temperatures T_e and cloud thicknesses l (and the corresponding value $S_\bullet$ appropriate for a given τ_e) and rejecting those values where $\tau_{ff} > 0.5\tau_e$ and $S_\bullet > 3.5$ Jy. We find that the reflection model can be satisfied for $T_e \geq 10^{6.7}$ K, $10^{6.2} \leq n_e < 10^{6.6}$ electrons cm^{-3} and $0.007 \leq l \leq 0.07$ pc (the upper limit set by the measured sizes). We also estimate that, assuming that thermal absorption is negligible, the flux density of any hidden compact radio source must be $0.8 \leq S_\bullet \leq 3.5$ Jy.

The second model is direct, thermal free-free emission from ionized gas inside the obscuring disk. Appropriate for the integrated radio spectrum^{9,14}, we assume for this thermal model a mean opacity of $\tau_{ff}(8.4 \text{ GHz}) = 0.5$ through the HZ plasma. Using the free-free opacity approximations of Mezger and Henderson¹⁵, we estimate $10^{6.5} \leq T_e \leq 10^{6.8}$ K and $n_e \geq 10^{6.8} \text{ cm}^{-3}$ ($T_e/10^7 \text{ K}$)^{1.35} ($l/0.07 \text{ pc}$)^{0.5}.

The plasma conditions in either the thermal or reflection models are plausible given the extreme environment. For comparison, particle densities in the molecular region of the disk are estimated to be $n_{\text{H}_2} \approx 10^8$ molecules cm^{-3} (refs 7, 16), and photoionization heating can drive T_e up to the limit bounded by inverse Compton cooling, $T_c \approx 10^7$ – 10^8 K (refs 7, 17). On the other hand, it is not clear how such a dense medium can remain heated to temperatures so near the Compton limit. For instance, photoionization heating of dense plasmas can support only $T_{\text{cool}} \approx 10^4$ – 10^6 K (refs 7, 18), unless the ionizing spectrum is much harder (more luminous in X-rays) and overall more luminous than current estimates¹⁹. A promising alternative is heating by mixing with a hot, intercloud medium²⁰, a $T_{\text{hot}} \approx 10^7$ – 10^8 K plasma proposed to confine optical emission-line clouds in the nuclear environment¹⁷. The temperatures in the mixed plasma would be $T_e \approx \sqrt{T_{\text{cool}} T_{\text{hot}}}$ (ref. 21), or $T_e \approx 10^{6.5}$ K, to within factors of a few. In addition, fast shocks, such as those driven by winds or the radio jet, or internal shocks arising at cloud–cloud collisions, might at least briefly support such high temperatures. Understanding the energy budget of the HZ will be a challenge for follow-up research.

Relevant specifically to AGN unifying schemes, the column density through the HZ clouds may be as high as $n_e l \approx 10^{24} \text{ cm}^{-2}$,

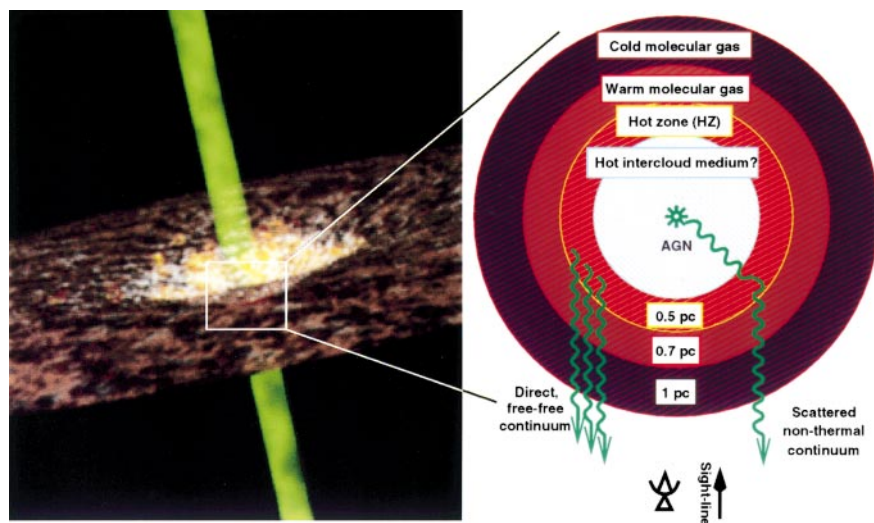


Figure 2 Diagrams of our model of the AGN of NGC1068 and its environs. Left panel, cartoon depicting the obscuring disk as viewed along our sight-line. The AGN is hidden from direct view, but we see reprocessed and scattered emission from the surrounding disk (illuminated material at the centre of the cartoon). Right panel, higher magnification plan view of the central part of our model (indicated by the white box in the left-hand panel). Shown are the relative locations of the 10^6 K 'hot zone' (HZ), detected in these observations, the warm transition zone, traced by H_2O maser emission and H I absorption¹⁰, and an outer, cooler molecular zone, which still eludes direct detection. It has been argued that the innermost region may be filled with a hot (10^8 K), intercloud medium¹⁷, which might be a source of heat for the HZ. This cartoon also illustrates two possible contributions to the observed radio emission: (1) scattered non-thermal emission originating at the AGN and (2) direct free-free emission from the HZ.

sufficient to absorb virtually all of the incident X-ray emission from the AGN²². This result, and the fact that the disk is viewed nearly edge-on, support the idea that the HZ traces ionized gas lying within the obscuring disk of NGC1068. One difficulty is that the covering fraction of the compact sources, $\sim 5\text{--}10\%$, is much smaller than required generally to explain the fraction of directly viewed AGNs²³ (the covering fraction is the portion of a sphere surrounding an AGN which is covered by obscuring clouds). However, the model can be reconciled if the geometric thickness of the obscuring medium increases with radius²⁴, and the HZ traces only emission from disk material nearest the AGN. Alternatively, the obscuring disk might instead be thin but highly warped²⁵, and again the HZ marks only the most central region of the disk.

Turning to the broad-band properties of the disk, the HZ plasma must also be a source of line and continuum emission up to soft X-ray energies. To determine whether the optical–X-ray spectrum of the HZ might be distinguished from neighbouring emission-line regions and the AGN proper, we modelled the HZ spectrum using the CLOUDY photoionization code²⁶, programmed to emulate a cooling plasma with the properties of the HZ, and normalized to the observed radio flux. We find that, in broad agreement with Pier and Voit⁷, the HZ contributes $\leq 10\%$ to the observed optical–ultraviolet emission lines of NGC1068 and $\leq 1\%$ to the optical–ultraviolet continuum. On the other hand, the HZ should contribute significantly to the soft X-ray spectrum of the nucleus were the AGN viewed along an unobscured sight-line. We estimate that free–free emission from the HZ may contribute $\geq 10\%$ of the AGN continuum in soft X-rays (photon energies $\sim 1\text{--}2\text{ keV}$). Moreover, the predicted soft X-ray line emission exceeds that observed towards NGC1068²⁷ by a factor of ≥ 100 . This observed diminution is approximately that expected for obscured X-ray emission viewed only in reflection¹⁹. Therefore, this result can be reconciled if the HZ is also heavily obscured over optical–soft X-ray wavebands. The implication is that, if the AGN unifying schemes hold generally, the disks surrounding unobscured AGN should be luminous soft X-ray emission-line sources. To our knowledge, there has as yet been no analysis of the soft X-ray line emission from unobscured AGN (that is, Seyfert 1 AGN). The detection of soft X-ray lines characteristic of a $10^6\text{--}10^7\text{ K}$ plasma in unobscured AGN would lend self-consistency to the obscuring disk model.

The present observation is, to our knowledge, the first direct image of a parsec-scale, ionized gas disk surrounding an AGN. Simple models for the radio emission further provide direct estimates of the physical properties of a parsec-scale disk, and the results are consistent with the predictions of AGN unifying schemes. However, we also find that photoionization heating is insufficient to support the high plasma temperatures in the disk, and so the challenge remains to model the energy budget in accord with the observed radio emission. Equally important is how the HZ might fit into the standard, infall model for AGN². The HZ is orientated nearly at right angles to the radio jet. The observed orientation suggests that, within the HZ, internal, viscous dissipation drives the fuelling of the AGN rather than external torques. In this regard, the HZ may be considered to define the outer extent of the long-sought accretion disk powering the AGN. □

Received 4 March; accepted 2 July 1997.

1. Rees, M. J. Black hole models for active galactic nuclei. *Annu. Rev. Astron. Astrophys.* **22**, 471–506 (1984).
2. Gunn, J. E. in *Active Galactic Nuclei* (eds Hazard, C. & Mitton, S.) 213–225 (Cambridge Univ. Press, 1979).
3. Begelman, M. C., Blandford, R. D. & Rees, M. J. Theory of extragalactic radio sources. *Rev. Mod. Phys.* **56**, 255–351 (1984).
4. Lawrence, A. Classification of active galaxies and the prospect of a unified phenomenology. *Publ. Astron. Soc. Pacif.* **99**, 309–334 (1987).
5. Antonucci, R. R. J. & Miller, J. S. Spectropolarimetry and the nature of NGC 1068. *Astrophys. J.* **297**, 621–532 (1985).
6. Krolik, J. H. & Begelman, M. C. Molecular tori in Seyfert galaxies—Feeding the monster and hiding it. *Astrophys. J.* **329**, 702–711 (1988).

7. Pier, E. A. & Voit, G. M. Photoevaporation of dusty clouds near active galactic nuclei. *Astrophys. J.* **450**, 628–637 (1995).
8. Gallimore, J. F., Baum, S. A. & O'Dea, C. P. The subarcsecond radio structure in NGC1068. II. Implications for the central engine and unifying schemes. *Astrophys. J.* **464**, 198–211 (1996).
9. Muxlow, T. W. B., Pedlar, A., Holloway, A. J., Gallimore, J. F. & Antonucci, R. R. J. The compact radio nucleus of the Seyfert galaxy NGC1068. *Mon. Not. R. Astron. Soc.* **278**, 854–860 (1996).
10. Gallimore, J. F., Baum, S. A., O'Dea, C. P., Brinks, E. & Pedlar, A. H₂O and OH masers as probes of the obscuring torus in NGC1068. *Astrophys. J.* **462**, 740–745 (1996).
11. Antonucci, R. R. J., Hurt, T. & Miller, J. S. HST ultraviolet spectropolarimetry of NGC1068. *Astrophys. J.* **430**, 210–217 (1994).
12. Kellerman, K. I. & Owen, F. N. in *Galactic and Extragalactic Radio Astronomy* (eds Verschuur, G. L. & Kellermann, K. I.) 563–600 (Springer, New York, 1989).
13. Capetti, A., Macchetto, F., Axon, D. J., Sparks, W. B. & Boksenberg, A. Hubble Space Telescope imaging polarimetry of the inner nuclear region of NGC 1068. *Astrophys. J.* **452**, L87–L89 (1995).
14. Gallimore, J. F., Baum, S. A., O'Dea, C. P. & Pedlar, A. The subarcsecond radio structure in NGC 1068. I. Observations and results. *Astrophys. J.* **458**, 136–148 (1996).
15. Metzger, P. G., Henderson, A. P. Galactic H II regions: I. Observations of their continuum radiation at the frequency 5 GHz. *Astrophys. J.* **147**, 471–489 (1967).
16. Neufeld, D. A., Maloney, P. R. & Conger, S. Water maser emission from X-ray-heated circumnuclear gas in active galaxies. *Astrophys. J.* **436**, L127–L130 (1994).
17. Krolik, J. H., McKee, C. F. & Tarter, C. B. Two-phase models for quasar emission line regions. *Astrophys. J.* **249**, 422–442 (1981).
18. Krolik, J. H. & Begelman, M. C. An X-ray heated wind in NGC 1068. *Astrophys. J.* **308**, L55–L58 (1986).
19. Pier, E. A., Antonucci, R., Hurt, T., Kriss, G. & Krolik, J. The intrinsic nuclear spectrum of NGC 1068. *Astrophys. J.* **428**, 124–129 (1994).
20. Reynolds, C. F. & Fabian, A. C. Warm absorbers in active galactic nuclei. *Mon. Not. R. Astron. Soc.* **273**, 1167–1176 (1995).
21. Begelman, M. C. & Fabian, A. C. Turbulent mixing layers in the interstellar and intracluster medium. *Mon. Not. R. Astron. Soc.* **244**, 26P–29P (1991).
22. Mulchaey, J. S., Mushotzky, R. F. & Weaver, K. A. Hard X-ray tests of the unified model for an ultraviolet-detected sample of Seyfert 2 galaxies. *Astrophys. J.* **390**, L69–L72 (1992).
23. Lawrence, A. The relative frequency of broad-lined and narrow-lined active galactic nuclei—Implications for unified schemes. *Mon. Not. R. Astron. Soc.* **252**, 586–592 (1991).
24. Efstathiou, A., Hough, J. H. & Young, S. A model for the infrared continuum spectrum of NGC 1068. *Mon. Not. R. Astron. Soc.* **277**, 1134–1144 (1995).
25. Sanders, D. B., Phinney, E. S., Neugebauer, G., Soifer, B. T. & Matthews, K. Continuum energy distribution of quasars—Shapes and origins. *Astrophys. J.* **357**, 29–51 (1989).
26. Ferland, G. *HAZY, a Brief Introduction to Cloudy* (Internal Rep., Dept of Physics and Astronomy, Univ. Kentucky, Lexington, 1993).
27. Ueno, S. et al. ASCA observations of NGC 1068. *Publ. Astron. Soc. Jpn* **46**, L71–L75 (1994).
28. Cornwell, T. & Braun, R. in *Synthesis Imaging in Radio Astronomy* (eds Perley, R. A., Schwab, F. R. & Bridle, A. H.) 167–181 (ASP Conf. Ser., Astron. Soc. Pacif., San Francisco, 1994).

Acknowledgements. The VLBA and VLA are operated by the National Radio Astronomy Observatory which is operated by Associated Universities, Inc., under cooperative agreement with the NSF. J.E.G. received support from a Collaborative Visitor's Grant from the Space Telescope Science Institute.

Correspondence and requests for materials should be addressed to J.E.G. (e-mail: jfg@hethp.mpe-garching.mpg.de).

Core formation on Mars and differentiated asteroids

Der-Chuen Lee and Alex N. Halliday

Department of Geological Sciences, University of Michigan, Ann Arbor, Michigan 48109-1063, USA

Meteorite chronometry based on the ^{182}Hf – ^{182}W system can provide powerful constraints on the timing of planetary accretion and differentiation^{1–4}, although the full potential of this method has yet to be realized. For example, no measurements have been made on the silicate-rich portions of planets and planetesimals other than the Earth and Moon. Here we report tungsten isotope compositions for two eucrites, thought to be derived from asteroid 4 Vesta, and from eight other basaltic achondritic meteorites that are widely considered to be from Mars. The eucrites, which are among the oldest differentiated meteorites, yield exceedingly radiogenic tungsten, indicating rapid accretion, differentiation and core formation on Vesta within the first 5–15 Myr of Solar System history, whereas the range of radiogenic tungsten measurements on the martian meteorites points towards tungsten depletion via melting and core formation within the first 30 Myr of the Solar System. The survival of tungsten isotope heterogeneity in the martian upper mantle implies that no giant impacts or large-scale convective mixing took place since this time. These results contrast with those obtained for the Earth–Moon system^{2,3} for which accretion and core formation related to

giant impacts appears to have continued for at least an additional 20 Myr.

Hafnium (Hf) and tungsten (W) are refractory elements whose relative proportions are strongly fractionated between silicate and metal. ^{182}Hf decays with a half-life of 9 Myr to ^{182}W such that the W isotope composition will be a function of the timing of Hf/W fractionation, and hence that of silicate-metal segregation^{1–4}. Despite a high Hf/W ratio, the silicate Earth has a W isotope composition identical to that of carbonaceous chondrites⁴, implying that the Earth's present core did not form until >50 Myr after the formation of the Solar System^{2,3} unless accretion was slower than generally modelled³. Iron meteorites and metal separates of ordinary chondrites contain W which is less radiogenic than chondritic, consistent with early segregation⁴. Tungsten which is more radiogenic than chondritic should be present in the silicate portions of some planets and planetesimals. Here we report results obtained on basaltic achondrites, thought to be derived from asteroid 4 Vesta and Mars. The siderophile element abundance patterns for both bodies are consistent with homogeneous accretion and core-mantle equilibration, possibly in a magma ocean⁵, in contrast to the non-equilibrium 'late veneer' model widely invoked to explain the excess siderophile concentrations in chondritic relative proportions that characterize the silicate Earth⁶.

Eucrites represent basaltic magmas erupted at, or near, the surface of the eucrite parent body (EPB), thought to be asteroid 4 Vesta (ref. 7). ^{60}Fe – ^{60}Ni , ^{53}Mn – ^{53}Cr and Pu–Xe data indicate that Vesta formed within the first ~20 Myr of the Solar System^{8–11}. The W isotope composition of two eucrites, ALHA78132 and Juvinas, are extremely radiogenic, corresponding to high, but different, Hf/W ratios (Table 1) and reflecting *in situ* ^{182}Hf decay within the EPB in overlapping stages: (1) accretion of the EPB with chondritic Hf/W; (2) metal segregation producing a silicate mantle with high Hf/W;

(3) melting to form a magma and rock with Hf/W slightly dissimilar from its source¹². A Hf–W model age (T_{CHUR}) can be calculated as follows:

$$T_{\text{CHUR}} = \frac{1}{\lambda} \ln \left\{ \left(\frac{^{182}\text{Hf}}{^{180}\text{Hf}} \right)_{\text{BSSI}} \times \left[\frac{\left(\frac{^{180}\text{Hf}}{^{184}\text{W}} \right)_{\text{sample}} - \left(\frac{^{180}\text{Hf}}{^{184}\text{W}} \right)_{\text{chond}}}{\left(\frac{^{182}\text{W}}{^{184}\text{W}} \right)_{\text{sample}} - \left(\frac{^{182}\text{W}}{^{184}\text{W}} \right)_{\text{chond}}} \right] \right\} \quad (1)$$

where λ is the decay constant of ^{182}Hf ($\sim 0.077 \text{ Myr}^{-1}$); $(^{182}\text{Hf}/^{180}\text{Hf})_{\text{BSSI}}$ is the initial isotopic composition of the Solar system, $(2.4 \pm 0.6) \times 10^{-4}$ (refs 2, 4); and chond indicates chondritic. This equation provides the age of metal segregation if Hf/W fractionation takes place in a single event. The two eucrites studied here yield similar Hf–W T_{CHUR} model ages in the range 5–15 Myr (Table 1), consistent with early metal segregation. More complex models such as concomitant core formation and accretion³ are also consistent with the W data, provided that the mean time for accretion is of the order of 3 Myr. No matter how complicated the model, the W isotope data appear inconsistent with the completion of metal segregation <4 Myr or >16 Myr after the start of the Solar System. This is in excellent agreement with the timescales inferred from ^{60}Fe – ^{60}Ni data (ref. 8), but is more difficult to reconcile with recent interpretations of ^{53}Mn – ^{53}Cr data (ref. 10). The absolute age of a eucrite can be determined by comparing its initial ^{53}Mn abundance with that of angrites, the most precisely dated of all meteorites¹³. On this basis the age of Juvinas has been estimated to be $4,562 \pm 1$ Myr, only 0.5–6.5 Myr later than the formation of the first condensates of the solar nebula¹⁴, just overlapping the timescale inferred from Hf–W. No independent age estimates are available for ALHA78132. However, the Cr isotope compositions of eucrites correlate with their respective Mn/Cr ratios in a manner interpreted to reflect primarily source heterogeneity inherited from planetary-scale partial melting at $4,565 \pm 1$ Myr (ref. 10), slightly earlier than allowed by the Hf–W model ages and inferred from ^{60}Fe – ^{60}Ni . Determination of detailed internal Hf–W isochrons should allow a direct comparison with Mn–Cr data and the resolution of this disparity.

SNC meteorites (shergottites, nakhlites and Chassigny) are young (<2 Gyr) achondrites which share chemical and oxygen isotope compositions consistent with a single large parent body, most likely to be Mars^{6,15}. An older meteorite, ALH84001, is different from the SNC meteorites, but seems to be from the same planet. ALH84001 and three shergottites (ALHA77005, Shergotty and Zagami) yield

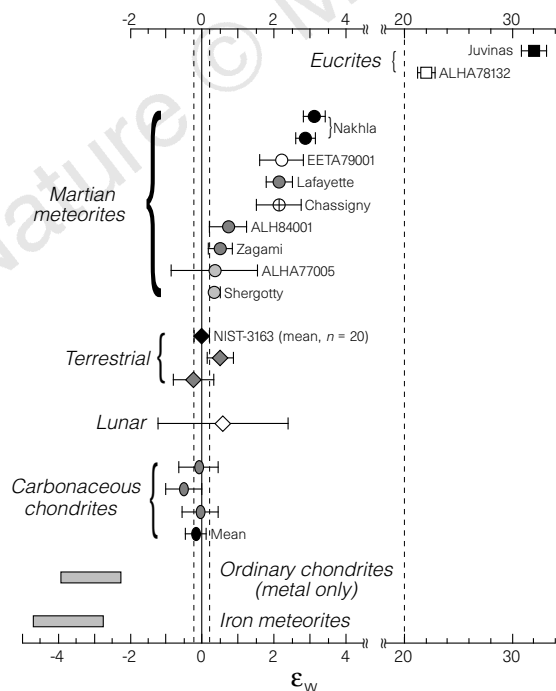


Figure 1 ϵ_{W} values of the two eucrites and eight martian meteorites analysed in this study (Table 1; $\epsilon_{\text{W}} = [(^{182}\text{W}/^{184}\text{W})_{\text{meas}} / (^{182}\text{W}/^{184}\text{W})_{\text{std}} - 1] \times 10^4$, relative to the NIST-3163 standard). Also plotted are previously reported data^{1–4} for comparison. Duplicates are shown with the same symbols. The ranges of ϵ_{W} values for the metal separates of ordinary chondrites and iron meteorites do not include the analytical uncertainties of individual samples.

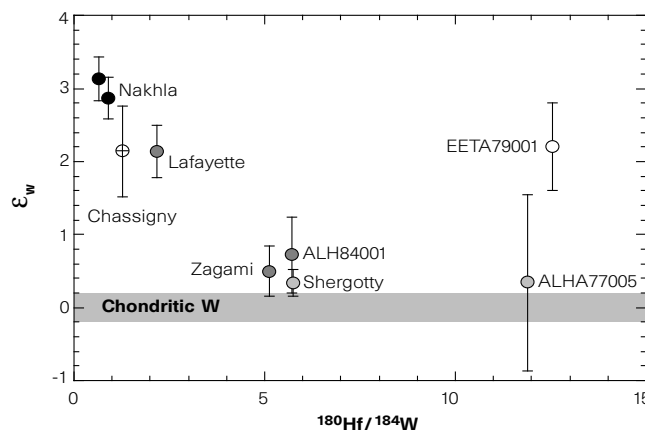


Figure 2 $^{180}\text{Hf}/^{184}\text{W}$ versus ϵ_{W} for eight martian meteorites measured in this study (Table 1), including a duplicate measurement of Nakhla. The horizontal shaded bar indicates the uncertainty over the W isotopic composition of carbonaceous chondrites.

Table 1 Hf and W isotope data for basaltic achondrites

Sample	Hf (p.p.m.)	W (p.p.m.)	$^{180}\text{Hf}/^{184}\text{W}$ (atomic)	$^{182}\text{W}/^{184}\text{W}$ $\pm 2\sigma$ error	ϵ_{W} $\pm 2\sigma$ error	T_{CHUR} (Myr)
Eucrites						
ALHA78132	1.014	0.07200	16.61	0.866901 ± 66	$+22.0 \pm 0.8$	8^{+3}_{-4}
Juvinas (USNM2844)	1.321	0.05606	27.81	0.86790 ± 17	$+33.5 \pm 2.0$	10^{+3}_{-4}
SNC meteorites						
ALH84001	0.2152	0.04449	5.707	0.865062 ± 45	$+0.72 \pm 0.52$	26^{+8}_{-23}
ALHA77005	0.5586	0.05543	11.89	0.86503 ± 10	$+0.34 \pm 1.2$	33^{+16}_{-33}
Shergotty (ME1372)	1.856	0.3804	5.756	0.865029 ± 16	$+0.34 \pm 0.28$	33^{+8}_{-24}
Zagami	1.946	0.4449	5.110	0.865043 ± 29	$+0.50 \pm 0.34$	29^{+18}_{-23}
EETA79001	0.8198	0.07709	12.55	0.865191 ± 52	$+2.21 \pm 0.60$	13^{+16}_{-13}
Lafayette (ME2116)	0.2801	0.1512	2.185	0.865185 ± 31	$+2.14 \pm 0.36$	13^{+10}_{-6}
Nakhla-1 (USNM5891)	0.2040	0.2641	0.9112	0.865248 ± 24	$+2.87 \pm 0.28$	10^{+5}_{-6}
Nakhla-2 (USNM5891)	0.1641	0.2976	0.6506	0.865271 ± 26	$+3.13 \pm 0.30$	9^{+5}_{-6}
Chassigny	0.04435	0.04106	1.275	0.865185 ± 54	$+2.14 \pm 0.62$	13^{+11}_{-11}

The differences in Hf and W concentrations in the two separately crushed aliquots of Nakhla are probably due to sample heterogeneity. The ϵ_{W} of each W isotope measurement is expressed as the deviations in parts per 10^4 relative to the NIST-3163 W standard, which gives a $^{182}\text{W}/^{184}\text{W} = 0.865000 \pm 18$ ($n = 20$). T_{CHUR} is defined in equation (1).

W which is close to chondritic ($\epsilon_{\text{W}} = 0$; see Fig. 1 for definition of ϵ_{W}), whereas two nakhrites (Nakhla and Lafayette), one shergottite (EETA79001) and Chassigny all display excess ^{182}W ($\epsilon_{\text{W}} = +2$ to $+3$; Table 1; Fig. 1).

ALH84001 is estimated to be >4 Gyr in age (ref. 16), but it must be younger than 4.53 Gyr given its chondritic W isotope composition. All other martian meteorites are relatively young igneous rocks^{17–20} and the measured Hf/W ratio, the complex product of a long history of melting, is clearly unrelated to the W isotope compositions which only relate to the earliest development of Mars (Fig. 2). It is, therefore, necessary to use the $^{180}\text{Hf}/^{184}\text{W}$ of 4 ± 1 for the primitive martian mantle (PMM) (refs 5, 21) in order to extract age information. This Hf/W ratio is much smaller than that estimated for the EPB⁵, so early core formation has not produced W which is as radiogenic as is found in the eucrites (Fig. 1). The samples with W close to chondritic yield imprecise model ages when calculated in this fashion (Table 1). However, those with radiogenic W yield Hf–W T_{CHUR} model ages indicative of early differentiation. The data for Nakhla indicate differentiation within the first 15 Myr of the Solar System. Modelling these data in terms of concomitant accretion and core formation³ yields a mean accretionary time of 5 Myr for Mars. The W isotope compositions could have been generated in an environment with higher Hf/W than the PMM, but the constraints

remain tight. Even if the Hf/W ratio were as high as EETA79001, three times the PMM, the data would still imply differentiation within the first 30 Myr.

These constraints from W isotopes are consistent with other isotope data for martian meteorites which also suggest early planetary-scale differentiation^{15–20,22}. Most importantly, the W data are in excellent agreement with data for ^{146}Sm – ^{142}Nd (half-life, 103 Myr) from the same martian meteorites²² (Fig. 3), also indicative of rapid accretion and differentiation. Those with excess ^{142}Nd , reflecting early depletion in light rare earths and other incompatible elements, are characterized by radiogenic W (Fig. 3). In order to generate a positive correlation between the W and Nd isotope data, the two parent/daughter ratios must have fractionated together at an early stage. If the core of Mars formed before the silicate melt depletion responsible for the high Sm/Nd, the enhanced Hf/W ratio of the silicate Mars would have driven the W isotope data to greater than chondritic values while the martian mantle maintained a chondritic Sm/Nd ratio and Nd isotope composition. This is not observed, as martian samples with chondritic Nd also have chondritic W isotope compositions (Fig. 3). Conversely, if core formation on Mars were later than silicate melt depletion, the presence of a metal phase during silicate melting would have buffered the concentrations of W and resulted in an inverse, or lack of, correlation between Sm/Nd and Hf/W in the melts. The positive correlation between Nd and W isotope data is also inconsistent with this model. The simplest explanation is that silicate melting and metal segregation (core formation) were coeval and co-genetic. Such a close relationship in a rapidly accreting planet is scarcely surprising as accretional energy and the decay of short-lived nuclides would have quickly heated the planet⁵, resulting in the formation of a shallow magma ocean from which metal segregated^{5,23}, as also indicated by metal–silicate equilibria for moderate siderophile elements⁵.

Although internal heating through the decay of U, Th and K will sustain melting, the occurrence of long-lived, large-scale vigorous mantle convection, as in the Earth, would effectively homogenize W isotope differences in the source regions of martian meteorites. Therefore, plate tectonics on Mars²⁴ would seem to be inconsistent with the W isotope data. This is consistent with the clear dichotomy between the ages of crust in the northern and southern hemispheres²⁵, the fact that ^{40}Ar is undegassed^{20,25}, a dynamically layered mantle²⁶ with perovskite at the base²⁷, and melting and tectonics driven by a single plume²⁸.

The $^{142}\text{Nd}/^{144}\text{Nd}$ and $^{182}\text{W}/^{184}\text{W}$ ratios thus far measured have been only chondritic or radiogenic, implying melt and metal depletion, albeit with subsequent partial re-enrichment. Although meteorite sampling may not be representative, the outer portions of Mars may be enriched in the cumulate residues of a magma ocean.

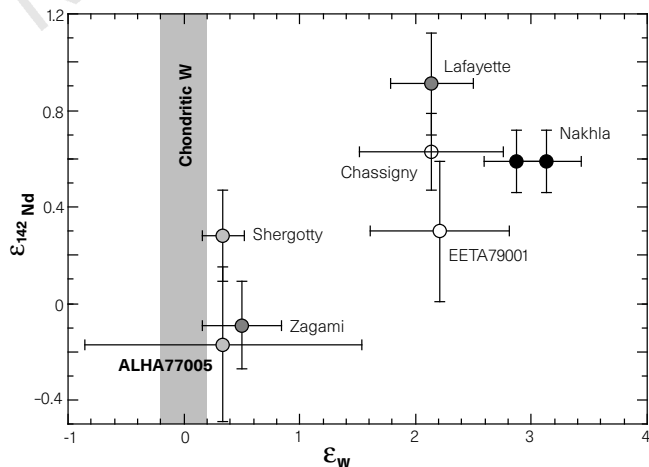


Figure 3 ϵ_{W} versus $\epsilon_{^{142}\text{Nd}}$ for the martian meteorites analysed in this study. The $\epsilon_{^{142}\text{Nd}}$ data are taken from Harper *et al.* (ref. 22). $\epsilon_{^{142}\text{Nd}}$ is the deviation in parts per 10^4 of the $^{142}\text{Nd}/^{144}\text{Nd}$ ratio from the terrestrial values. The vertical shaded bar indicates the uncertainty for the W isotopic composition of carbonaceous chondrites.

This would involve very dense Fe-rich melts¹⁵, particularly before the completion of core formation, and some of these melts may have partially sunk with core-forming metal towards the martian interior. Cumulate and residue flotation depends on the density crossover between solid and liquid phases²⁷, which is poorly constrained for Mars. However, if the outer portions of the planet were rich in depleted residual solids during the magma ocean stage, the W and Nd isotope compositions would be expected to be radiogenic and to survive homogenization.

The sizes of inner Solar System planetary bodies seem to increase with accretion and differentiation intervals, from the most primitive and undifferentiated materials such as equilibrated chondrites (~6 Myr) to the eucrite parent body (~10 Myr) to Mars (10–30 Myr) and to the Earth–Moon system (>50 Myr) (refs 2, 3). The data are consistent with early termination of accretion of asteroids and Mars, and longer-lived accretion of larger bodies resulting from late planet-scale impacts. The trend corresponds to a transition in the dominant mechanisms of heating from decay of short-lived nuclides to the release of accretional energy. Whereas a late giant impact on Earth may have been sufficiently energetic to effectively homogenize the W isotope composition and destroy any proto-core^{2,3}, the W and Nd isotope data are inconsistent with an impact sufficiently energetic to melt and homogenize Mars later than 4.53 Gyr (ref. 22). □

Methods. All samples were powdered in an aluminum oxide mortar after surface saw marks and fusion crusts had been removed through mild acid leaching (1M HCl) and hand-picking. The samples were digested sequentially with concentrated HF, 8M HNO₃ and 6M HCl. Roughly 10% of the solution was separated and spiked with ¹⁷⁸Hf and ¹⁸⁶W, whereas the remaining solution was dried and redissolved in ~8 ml of 4M HF. The method of chemical separation of W was adapted from the first column of the Hf chemistry developed by Salters and Hart²⁹ but on a reduced scale, with 3.5 ml of Bio-Rad AG1 × 8 (200–400 mesh) anion resin. The Hf and W were eluted and collected sequentially using a mixed solution of 6M HCl + 1M HF. The same chemical procedure was used for the spiked solutions, but the column volume was further reduced (~1 ml), and Hf and W were collected together. The total W procedural blank was ~0.4 ng. Tungsten isotopic measurement by multiple collector inductively coupled-plasma mass spectrometry (MC-ICPMS) has been previously described by Lee and Halliday³⁰. The NIST-3163 W standard was run between every sample to monitor the performance of MC-ICPMS and to check for memory effects, which were negligible. The W isotopic measurements were normalized to ¹⁸⁶W/¹⁸⁴W = 0.927633 (ref. 30). The quoted 2σ standard errors all refer to the least significant figures. The concentrations of Hf and W were determined by isotope dilution, and the uncertainty is typically 0.2% or better. All the W data are blank-corrected, which is negligible for all analyses except for Juvinas which has been corrected by 1.5 ε unit.

Received 3 April; accepted 15 July 1997.

- Harper, C. L. & Jacobsen, S. B. Evidence for ¹⁸²Hf in the early Solar System and constraints on the timescale of terrestrial accretion and core formation. *Geochim. Cosmochim. Acta* **60**, 1131–1153 (1996).
- Lee, D.-C. & Halliday, A. N. Hafnium–tungsten chronometry and the timing of terrestrial core formation. *Nature* **378**, 771–774 (1995).
- Halliday, A. N., Rehkämper, M., Lee, D.-C. & Yi, W. Early evolution of the Earth and Moon: new constraints from Hf–W isotope geochemistry. *Earth Planet. Sci. Lett.* **142**, 75–89 (1996).
- Lee, D.-C. & Halliday, A. N. Hf–W isotopic evidence for rapid accretion and differentiation in the early solar system. *Science* **274**, 1876–1879 (1996).
- Righter, K. & Drake, M. J. Core formation in Earth's Moon, Mars and Vesta. *Icarus* **124**, 513–529 (1996).
- Wänke, H. & Dreibus, G. Chemical composition and accretion history of terrestrial planets. *Phil. Trans. R. Soc. Lond. A* **325**, 545–557 (1988).
- Binzel, R. P. & Xu, S. Chips off of asteroid 4 Vesta: evidence for the parent body of basaltic achondrite meteorites. *Science* **260**, 188–191 (1993).
- Shukolyukov, A. & Lugmair, G. W. ⁶⁰Fe in eucrites. *Earth Planet. Sci. Lett.* **119**, 159–166 (1993).
- Lugmair, G. W. & MacIsaac, G. Radial heterogeneity of ⁵³Mn in the early solar system? *Lunar Planet. Sci.* **XXVI**, 879–880 (1995).
- Lugmair, G. W. & Shukolyukov, A. ⁵³Mn–⁵³Cr isotope systematics of the HED parent body. *Lunar Planet. Sci.* **XXVIII**, 651–852 (1997).
- Shukolyukov, A. & Begemann, F. Pu–Xe dating of eucrites. *Geochim. Cosmochim. Acta* **60**, 2453–2471 (1996).
- Newsom, H. E., White, W. M., Jochum, K. P. & Hofmann, A. W. Siderophile and chalcophile element abundances in oceanic basalts, Pb isotope evolution and growth of the earth's core. *Earth Planet. Sci. Lett.* **88**, 299–313 (1986).

- Lugmair, G. W. & Galer, S. J. G. Age and isotopic relationships among the angrites Lewis Cliff 86010 and Angr dos Reis. *Geochim. Cosmochim. Acta* **56**, 1673–1694 (1992).
- Göpel, C., Manhès, G. & Allegre, C. J. Constraints on the time of accretion and thermal evolution of chondrite parent bodies by precise U–Pb dating of phosphates. *Meteorites* **26**, 73 (1991).
- McSween, H. Y. What we have learned about Mars from SNC meteorites. *Meteoritics* **29**, 757–779 (1994).
- Jagoutz, E., Sorowka, A., Vogel, J. D. & Wänke, H. ALH84001: alien or progenitor of the SNC family? *Meteoritics* **9**, 478–479 (1994).
- Nakamura, N., Unruh, D. M. & Tatsumoto, M. Origin and evolution of the Nakhla meteorite inferred from the Sm–Nd and U–Pb systematics and REE, Ba, Sr, Rb and K abundances. *Geochim. Cosmochim. Acta* **46**, 1555–1573 (1982).
- Shih, C.-Y. et al. Chronology and petrogenesis of young achondrites, Shergotty, Zagami, and ALHA77005: late magmatism on a geologically active planet. *Geochim. Cosmochim. Acta* **46**, 2323–2344 (1982).
- Chen, J. H. & Wasserburg, G. J. Formation ages and evolution of Shergotty and its parent planet from U–Th–Pb systematics. *Geochim. Cosmochim. Acta* **50**, 955–968 (1986).
- Jagoutz, E. & Wänke, H. Sr and Nd isotopic systematics of Shergotty meteorites. *Geochim. Cosmochim. Acta* **50**, 939–953 (1986).
- Treiman, A. H. et al. Core formation in the Earth and shergottite parent body (SPB): chemical evidence from basalts. *Geochim. Cosmochim. Acta* **50**, 1071–1091 (1986).
- Harper, C. L. et al. Rapid accretion and early differentiation of Mars indicated by ¹⁴²Nd/¹⁴⁴Nd in SNC meteorites. *Science* **267**, 213–217 (1995).
- Stevenson, D. J. in *Origin of the Earth* (eds Newsom, H. E. & Jones, J. H.) 231–249 (Oxford Univ. Press, London, 1990).
- Sleep, N. H. Martian plate tectonics. *J. Geophys. Res.* **99**, 5639–5655 (1994).
- Breuer, D., Spohn, T. & Wüllner, U. Mantle differentiation and the crustal dichotomy of Mars. *Planet. Space Sci.* **41**, 269–283 (1993).
- Breuer, D., Yuen, D. A. & Spohn, T. Phase transitions in the Martian mantle: implications for partially layered convection. *Earth Planet. Sci. Lett.* **148**, 457–469 (1997).
- Bertka, C. M. & Fei, Y. Mineralogy of the Martian interior up to core-mantle boundary pressures. *J. Geophys. Res.* **102**, 5251–5264 (1997).
- Harder, H. & Christensen, U. R. A one-plume model of martian mantle convection. *Nature* **380**, 507–509 (1996).
- Salters, V. J. M. & Hart, S. R. The mantle sources of ocean ridges, islands and arcs: the Hf-isotope connection. *Earth Planet. Sci. Lett.* **104**, 364–380 (1994).
- Lee, D.-C. & Halliday, A. N. Precise determinations of the isotopic compositions and atomic weights of molybdenum, tellurium, tin and tungsten using ICP magnetic sector multiple collector mass spectrometry. *Int. J. Mass Spec. Ion Proc.* **146/147**, 35–46 (1995).

Acknowledgements. We thank M. Lindstrom, L. Nyquist, G. MacPherson, C. Perron and M. Wadhwa for access to their meteorite collections at NASA, Smithsonian Institution of Washington, Museum National d'Histoire Naturelle at Paris, and Field Museum in Chicago. We also thank J. Christensen, E. Essene, H. Pollack, M. Rehkämper, P. van Keken and Y. Zhang for their comments, M. Johnson and C. Hall for their assistance, and K. Righter and M. Drake for access to unpublished papers. This work was supported by NSF, DOE, NASA and the University of Michigan.

Correspondence and requests for materials should be addressed to D.-C.L. (e-mail: dcllee@umich.edu)

Effect of microgravity on the crystallization of a self-assembling layered material

Homayoun Ahari*, Robert L. Bedard†, Carol L. Bowes*, Neil Coombs‡, Ömer Dag*, Tong Jiang*, Geoffrey A. Ozin*, Srebri Petrov*, Igor Sokolov*, Atul Verma*, Gregory Vovk* & David Young*

* Materials Chemistry Research Group, Lash Miller Chemical Laboratories, University of Toronto, Toronto, Ontario M5S 3H6, Canada

† UOP, Research Division, 25 E. Algonquin Road, Des Plaines, Illinois 60017, USA

‡ Imagetek Analytical Imaging, 32 Manning Avenue, Toronto, Ontario M6J 2K4, Canada

In microgravity, crystals of semiconductors and proteins can be grown with improved crystallinity, offering the prospect of improved structural analyses (for proteins) and better electronic properties (for semiconductors)^{1–3}. Here we study the effect of a microgravity environment on the crystallization of a class of materials—layered microporous tin(IV) sulphides^{4–11}—whose crystal structure is determined by weak interlayer interactions (electrostatic, hydrogen-bonding and van der Waals) as well as strong intralayer covalent bonds. We find that the crystals grown in microgravity (on board the Space Shuttle Endeavour) show improved crystal habits, smoother faces, greater crystallinity, better optical quality and larger void volumes than the materials grown on Earth. These differences are due at least in part to the

profound influence of microgravity on the layer registry over length scales of around a nanometre, which is shown by X-ray and electron diffraction to be better in space than on Earth. Thus we can see a clear distinction between the covalent bonds in these materials, which are not significantly affected by microgravity, and the weaker forces (like those that determine the structure of proteins over length scales of around 0.3–0.4 nm) which are more susceptible to the dynamic disturbances that operate in crystallization on Earth.

The use of microgravity in our studies of layered microporous tin(IV) sulphides was stimulated by the sensitivity of their crystal growth to the uniformity of the solution environment and the extreme difficulty of growing diffraction-quality single crystals. It was anticipated that a comparative study of the morphology, surface topology and microstructure of crystals prepared in space and on Earth would provide an insight into the nucleation and growth steps that are affected by gravity-driven convective mass transport and Stokes sedimentation effects. This knowledge could lead to methods for producing better crystals in Earth-based processes.

A number of layered microporous tin(IV) sulphides have recently been synthesized and structurally characterized (Fig. 1 top)^{4–11}. The current understanding of their mode of formation is shown in Fig. 1 (bottom)^{9,10}. In brief, solution-phase ($\text{Sn}_2\text{S}_6^{4-}$) building units 'snap' together, as donor–acceptor pairs, to create Sn_3S_4 broken-cube clusters. These assemble around organic template cations R^+ (where R^+ is TMA^+ ($\text{N}(\text{CH}_3)_4^+$) or TBA^+ ($\text{N}(\text{C}_4\text{H}_9)_4^+$)) to form porous anionic $[\text{Sn}_n\text{S}_{2n+1}^{2-}]$ sheets. The extra sulphur is eliminated as sulphide. The templates direct the sheets into poorly organized stacks which move into registry in the $\text{R}_2\text{Sn}_n\text{S}_{2n+1}$ product.

To study the effect of microgravity on this process, a total of 19 syntheses were selected based on the results of two years of ground-based study involving variations of reaction stoichiometry, choice of precursors, heating profile, co-solvents and ageing. Synthesis of the orthorhombic $\text{TBA}_2\text{Sn}_4\text{S}_9$ (space group *Pbcn*) and $\text{TMA}_2\text{Sn}_3\text{S}_7$ ($P2_12_12_1$) were optimized using the precursors $x\text{ROH}:y\text{S}:z\text{SnS}_2:p\text{H}_2\text{O}:q\text{TEG}$, where R^+ is the organic template cation and TEG is the co-solvent tetraethyleneglycol. Each

microgravity preparation was run in duplicate for a total of 38 reaction mixtures to check for the reproducibility of the results. Every one of these was repeated, under identical conditions, for the Earth-based study. The reaction mixtures were aged in their crucibles for a period of 79 days during which time they were transferred to the Kennedy Space Center, installed in Endeavour, lifted off into space at 5:30 on 19 May 1996 and heated to reaction temperature on 22 May 1996. The temperature of each tube was independently controlled with a computer which recorded the temperature profile. Samples were heated to 150 °C over 10 hours, maintained at this temperature for 13 hours 23 minutes, and cooled to 25 °C over 24 hours. The space samples were aged for 42 days during which time they were transferred from the Space Shuttle to our laboratory. A special lathe removed the steel caps from the containers for the crucible recovery process. All space products were filtered, washed with water and acetone, air dried and stored under argon in the dark. The reaction conditions and ageing steps were reproduced to the best of our ability for the Earth samples.

Representative scanning electron microscopy (SEM) images of space- and Earth-grown samples of $\text{TBA}_2\text{Sn}_4\text{S}_9$ are shown in Fig. 2 left. The corresponding powder X-ray diffraction (PXRD) patterns are displayed in Fig. 2 right. Crystals grown in space have well-formed rectangular plate-like habits. This is the expected morphology for the orthorhombic *Pbcn* space group and layered structure of the material. The dimensions of the {102} face of the crystals fall in the range 30–40 µm. The SEM images of the Earth-grown crystals reveal significant differences in morphology and size. They have submicrometre dimensions and consist of spherical agglomerates. The PXRD pattern for the microgravity sample displays better-resolved reflections which indicate a much higher degree of layer registry and overall crystallinity. Similar effects are observed in the SEM images and PXRD patterns of $\text{TMA}_2\text{Sn}_3\text{S}_7$ grown in space and on Earth. The full-width at half-maximum values of the 0k0 reflections that correspond to the interlayer spacing are consistently smaller for the space-grown crystals, demonstrating that they have better layer registry than the Earth analogues. The spacing between

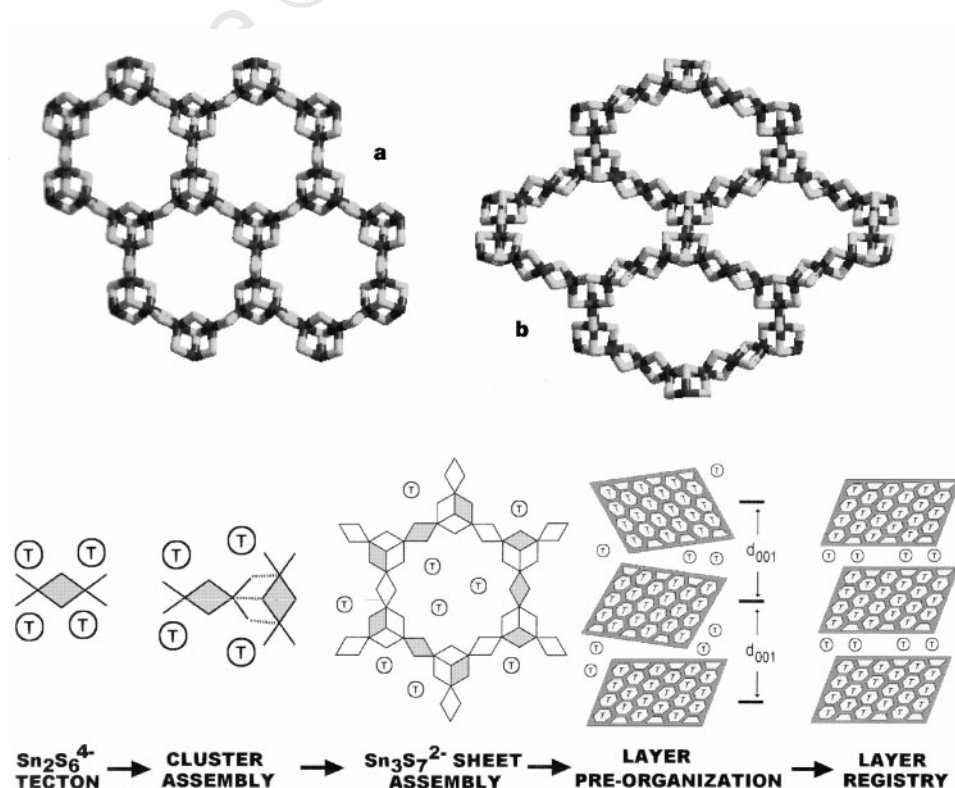


Figure 1 Top, the structure of the microporous sheets of $\text{TMA}_2\text{Sn}_3\text{S}_7$ (a) and $\text{TBA}_2\text{Sn}_4\text{S}_9$ (b). The template cations are not shown. Bottom, current understanding of the model of formation of layered microporous tin(IV) sulphides, showing precursors, intermediates and products; modular construction from dimer building-blocks to layer registry is shown (here T stands for template)^{9–11}.

the layers in these materials is ~ 9 Å for $\text{TMA}_2\text{Sn}_3\text{S}_7$ and ~ 14 Å for $\text{TBA}_2\text{Sn}_4\text{S}_9$ (ref. 11). The anionic layers are held together by weak electrostatic, hydrogen-bonding and van der Waals interactions with the template cations. Organization of the layers into registered stacks is expected to be a low-activation-energy process and sensitive to the uniformity of the solution environment. By contrast, the forces governing ordering within the layers involve strong covalent bonds which are less likely to be perturbed by the growth environment (see below). Note that co-existing spherical particles seen in the SEM images of the Earth-grown samples are unreacted tin (see also PXRD) and are less prevalent in the space samples. Tin particles tend to sediment in a 1g environment and passivation by sulphur can occur. In microgravity, tin particles will instead levitate in solution with less opportunity to passivate and leave unreacted tin behind.

Representative electron diffraction (ED) patterns of space- and Earth-grown crystals of $\text{TMA}_2\text{Sn}_3\text{S}_7$ are shown in Fig. 3. The beam direction was arranged in both cases to be orthogonal to the dominant $\{010\}$ face of the hexagonal-shaped crystals. Under similar recording conditions (that is, size and thickness of crystals, beam location on crystals, aperture size and exposure time) the space-grown crystals consistently show higher-intensity ED patterns with no streaking of spots compared with the Earth analogues. The ED pattern of the space crystal corresponds to that expected for the orthorhombic $P2_12_12_1$ space group of $\text{TMA}_2\text{Sn}_3\text{S}_7$. This ED pattern is to be contrasted with the one observed for the Earth-grown crystal where the streaking between ED spots probably originates from interlayer stacking faults (planar defects)¹². This difference implies that gravity contributes to the creation of inter-

layer disorder. Overall, the ED results show better crystallinity and registry of the layers for the space-grown crystals.

The adsorption of CO_2 at 77 K by $\text{TMA}_2\text{Sn}_3\text{S}_7$ shows a type I isotherm (diagnostic of a microporous material) for both space and Earth samples. A Langmuir plot of the data shows that the accessible void volume determined for the space sample ($0.045 \text{ cm}^3 \text{ g}^{-1}$) is $\sim 60\%$ greater than that of the Earth analogue. Better-aligned porous layers are expected to provide less restricted channels running between the sheets. As a result, the CO_2 adsorbed into the space-grown material experiences less tortuous diffusion pathways into the structure and has a larger available void volume than the Earth sample.

Atomic-force microscopy (AFM) has been used to quantify the roughness of crystal surfaces over an area of $1 \times 1 \mu\text{m}^2$. Averaging has been conducted for five crystal samples with ten area measurements for each. The space-grown crystals consistently have smoother faces (roughness of 5–30 Å) than the Earth-grown analogues (25–100 Å). By contrast, there is little difference between the degree and spatial extent of surface microstructural order in these samples. The surface parallel to the $\{010\}$ face of space- and Earth-grown $\text{TMA}_2\text{Sn}_3\text{S}_7$ crystals shows a well resolved orthorhombic pattern with high-contrast features separated by a distance of 13 ± 1.0 Å. This corresponds to the spacing between capping-sulphur atoms of alternate Sn_3S_4 broken-cube clusters arranged in an up-down configuration around the 24-atom pore (Fig. 1). Similarly, AFM images of the dominant surface parallel to the $\{102\}$ face of space and Earth $\text{TBA}_2\text{Sn}_4\text{S}_9$ crystals reveal a well-defined orthorhombic pattern. High-contrast features with a separation of 16.0 ± 1.0 Å and 19.2 ± 1.2 Å correspond to the

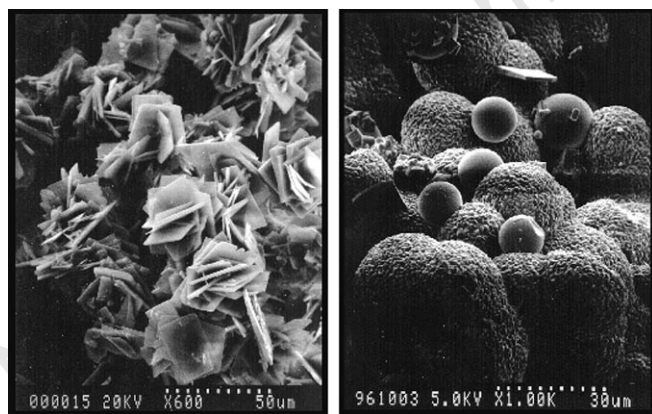
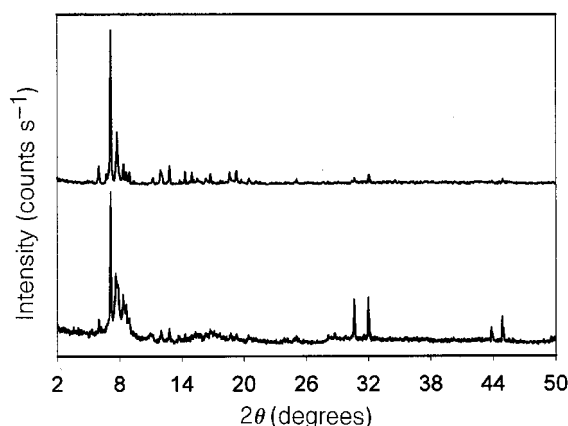


Figure 2 Above, scanning electron microscopy images of $\text{TBA}_2\text{Sn}_4\text{S}_9$ grown in space (left; scale bar $50 \mu\text{m}$) and on Earth (right; scale bar $30 \mu\text{m}$); far right, powder



X-ray diffraction patterns of $\text{TBA}_2\text{Sn}_4\text{S}_9$ grown in space (top trace) and on Earth (bottom trace). Reflections around $30\text{--}32^\circ$ and $44\text{--}46^\circ$, 2θ depict elemental tin.

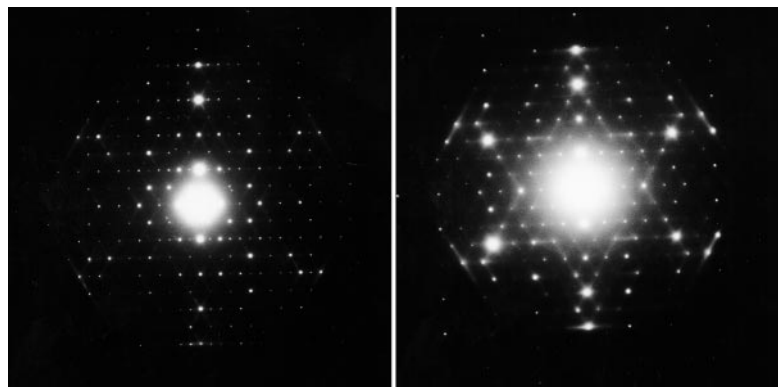


Figure 3 Electron diffraction patterns of orthorhombic $P2_12_12_1$ $\text{TMA}_2\text{Sn}_3\text{S}_7$ grown in space (left) and on Earth (right).

spacings between the capping-sulphur atoms of alternate Sn_3S_4 broken-cube clusters that are arranged in an up-down configuration around the 32-atom pore. From these AFM results it can be deduced that the less turbulent solution environment in microgravity enhances the mesoscopic smoothness of the dominant crystal face that contains the layer planes in $\text{TMA}_2\text{Sn}_3\text{S}_7$ and $\text{TBA}_2\text{Sn}_4\text{S}_9$. However, the microscopic order within the layers is about the same for the space- and Earth-grown materials (see below).

Polarizing optical microscopy images for $\text{TMA}_2\text{Sn}_3\text{S}_7$ space- and Earth-grown crystal samples have been recorded. The uniform and complete extinction of white light observed between crossed polarizers for the space-grown samples confirms that the space samples are single crystals. By contrast, the Earth-grown samples display mosaic patterns and birefringence interference colours in their images; these patterns and colours are diagnostic of crystals consisting of multicrystalline agglomerations (this has been confirmed by laser scanning confocal optical microscopy). The difference could originate from increased populations of defects in the Earth-grown samples which serve as secondary nucleation sites for the formation of crystal agglomerates. Photoluminescence studies show that the space samples do indeed contain fewer defects.

Optical reflectance spectroscopy has revealed a pronounced enhancement of intensity on the optical absorption edge but without a shift of the bandgap for Earth-grown compared with space-grown $\text{TMA}_2\text{Sn}_3\text{S}_7$ samples. This difference probably has its origin in a relaxation of the selection rules for $\text{S}(-\text{II})$ to $\text{Sn}(\text{IV})$ interband electronic transitions for the Earth samples because of poorer registry of the layers.

The techniques of ^{119}Sn magic-angle-spinning NMR, ^{119}Sn Mössbauer spectroscopy and Fourier-transform Raman spectroscopy proved to be effective probes of local electronic and geometrical structure around the tin(IV) sites in $\text{TMA}_2\text{Sn}_3\text{S}_7$ and $\text{TBA}_2\text{Sn}_4\text{S}_9$. No difference was found in the following parameters between space- and Earth-grown samples: NMR isotropic chemical shifts, Mössbauer isomer shifts and quadrupole splittings, Raman vibrational frequencies, intensities and spectral line shapes. The electrical response of these materials to the adsorption of guest molecules has been investigated by an *in situ* conductivity study. The resistance of $\text{TMA}_2\text{Sn}_3\text{S}_7$ drops drastically, up to 4–5 orders of magnitude, when the sample *in vacuo* is exposed to water at 20 °C. The electrical conductivities as a function of partial pressure of the adsorbate for the space-grown and Earth-grown samples are, however, virtually indistinguishable. Because the conductivity of $\text{TMA}_2\text{Sn}_3\text{S}_7$ is expected to be highly anisotropic with greater intralayer than interlayer charge transport, the similarity of the conductivity isotherms of the space- and Earth-grown samples again implies that intralayer order is not significantly influenced by gravity. Thus different analytical techniques can reveal the steps in the crystallization of a self-assembling microporous layered material that are under the influence of gravity.

Although the results of these experiments have not established that gravity alone is the controlling factor in these experiments, there is evidence that the observed differences in interlayer registry and population of defects between space- and Earth-grown crystals probably originate in a microgravity effect rather than a change in the kinetics of the process. Full details will appear elsewhere¹³. □

Received 19 February; accepted 25 June 1997.

1. Feuerbacher, B., Hamacher, H. & Naumann, R. J. (eds) *Materials Sciences in Space* (Springer, Berlin, 1986).
2. *Space Science in the Twenty-First Century. Imperatives for the Decades 1995–2015—Overview and Fundamental Physics and Chemistry* (Space Science Board, Natl Res. Council, Natl Academy Press, Washington DC, 1988).
3. *Towards a Microgravity Research Strategy* (Committee on Microgravity Research, Space Studies Board, Natl Academy Press, Washington DC, 1992).
4. Ahari, H. *et al.* Microporous tin(IV) chalcogenides: flexible framework materials for molecule discrimination. *Adv. Mater.* **7**, 375–378 (1995).
5. Jiang, T. *et al.* Synthesis and structures of novel microporous tin(IV) sulfide materials. TPA-SnS-3 and TBA-SnS-3. *Chem. Mater.* **7**, 245–248 (1995).

6. Ozin, G. A. & Bowes, C. L. Self-assembling frameworks. *Adv. Mater.* **8**, 13–28 (1996).
7. Ozin, G. A. Microporous and mesoporous electronic materials: towards the electronic nose. *Supramol. Chem.* **6**, 125–134 (1995).
8. Ozin, G. A. *et al.* Electronic sieves: Microporous layered tin(IV) sulfides. A new class of sensory materials. *Proc. 9th Int. Symp. on Molecular Recognition and Inclusion* (ed. Coleman, A.) (Kluwer Academic, Dordrecht, Netherlands, 1997).
9. Jiang, T., Ozin, G. A. & Young, D. Microporous tin(IV) sulfides: mode of formation. *Adv. Mater.* **6**, 860–865 (1994).
10. Francis, R. J. *et al.* Hydrothermal synthesis of microporous tin sulfides studied by real time *in situ* energy dispersive X-ray diffraction. *Chem. Mater.* **8**, 2102–2108 (1996).
11. Jiang, T. *Porous Tin(IV) sulfides*, Thesis, Univ. Toronto, 1997.
12. Williams, D. B. & Carter, C. B. *Transmission Electron Microscopy. A Textbook for Materials Science* (Plenum, New York, 1996).
13. Dag, Ö. *et al.* Does microgravity influence self-assembly? A comparative investigation of the crystal growth of self-assembling microporous layered tin(IV) sulfides under microgravity conditions. *Adv. Mater.* (in the press).

Acknowledgements. This work was supported by the Natural Sciences and Engineering Research Council of Canada, Universal Oil Products, Communication Development, the Canadian Space Agency and NASA.

Correspondence and requests for materials should be addressed to G.A.O. (e-mail: gozin@alchemy.chem.utoronto.ca).

Biodegradable block copolymers as injectable drug-delivery systems

Byeongmoon Jeong, You Han Bae, Doo Sung Lee & Sung Wan Kim

Biomedical Polymers Research Building, Room 205, CCCD/Pharmaceutics, University of Utah, Salt Lake City, UT 84112, USA

Polymers that display a physicochemical response to stimuli are widely explored as potential drug-delivery systems^{1–4}. Stimuli studied to date include chemical substances and changes in temperature, pH and electric field. Homopolymers or copolymers of *N*-isopropylacrylamide^{5,6} and poly(ethylene oxide)–poly(propylene oxide)–poly(ethylene oxide) (known as poloxamers)⁷ are typical examples of thermosensitive polymers, but their use in drug delivery is problematic because they are toxic and non-biodegradable. Biodegradable polymers used for drug delivery to date have mostly been in the form of injectable microspheres or implant systems, which require complicated fabrication processes using organic solvents⁸. Such systems have the disadvantage that the use of organic solvents can cause denaturation when protein drugs are to be encapsulated. Furthermore, the solid form requires surgical insertion, which often results in tissue irritation and damage. Here we report the synthesis of a thermosensitive, biodegradable hydrogel consisting of blocks of poly(ethylene oxide) and poly(L-lactic acid). Aqueous solutions of these copolymers exhibit temperature-dependent reversible gel–sol transitions. The hydrogel can be loaded with bioactive molecules in an aqueous phase at an elevated temperature (around 45 °C), where they form a sol. In this form, the polymer is injectable. On subcutaneous injection and subsequent rapid cooling to body temperature, the loaded copolymer forms a gel that can act as a sustained-release matrix for drugs.

Diblock copolymers consisting of poly(ethylene oxide) (PEO) and poly(L-lactic acid) (PLLA) were synthesized by ring opening polymerization of L-lactide (LLA) on monomethoxy poly(ethylene oxide); triblock copolymers were synthesized by coupling the diblock copolymers using hexamethylene diisocyanate (HMDI) as a coupling agent (Fig. 1). It is known that PEO of high relative molecular mass (M_r above ~10,000) is unsuitable for filtration through human kidney membrane because of the large hydrodynamic radius of the PEO in an aqueous phase. Therefore we

used PEO segments of M_r 5,000. The incorporation of a biodegradable block into the polymer backbone will result in the eventual decrease in the polymer M_r after injection, leading to easier elimination from the body than is the case for non-biodegradable polymers (such as poloxamers).

Aqueous solutions of both diblock (PEO–PLLA) and triblock (PEO–PLLA–PEO) copolymers form micelles at low concentrations. This was confirmed by dye solubility experiments during which we observed an abrupt enhancement in the ultraviolet (356-nm) absorption of the dye diphenylhexatriene (DPH). The enhanced absorption reflects the sudden increase in dye molecules that are dissolved by incorporation in micelles, with the formation of micelles occurring once the copolymer concentration exceeds a critical value⁹. The micelle formation of the block copolymers is driven by hydrophobic interactions, that is, by packing of the hydrophobic PLLA blocks. At high concentration levels, the aqueous polymer solutions form a gel, with this gelation probably being caused by the association of micelles. These block copolymers formed a gel at a lower temperature and became a sol at a higher temperature. By changing the biodegradable block length, the sol–gel transition temperature can easily be manipulated (Fig. 2); increasing the PLLA block length increases the aggregation tendency of a block copolymer in water, resulting in a steepening of the gel–sol transition curve slopes and the onset of gelation at lower concentration. The sol–gel transition temperature is a function of concentration as well as composition of a block copolymer.

The system is designed mainly for delivery of high- M_r protein drugs which have a low diffusion coefficient. An *in vitro* release study was conducted for fluorescein isothiocyanate (FITC) labelled dextran (M_r 20,000) as a model drug. The system can also be used for sustained release of low- M_r hydrophobic drugs. The sol–gel transition temperature is affected by the presence of a drug. These polymers and the bioactive agents are dissolved in water above body temperature (45°C). FITC-labelled dextran was mixed with the polymer solution at 45°C, followed by lowering the temperature to 37°C. Afterwards, the solutes released from the formed gel were monitored by fluorescence spectroscopy. The release of FITC-labelled dextran from the gel is shown as a function of time in Fig. 3. The release rate can be controlled by initial loading, the M_r or hydrophobicity of the drug, and the concentration of polymer. With an increase in polymer concentration, release rates decrease (Fig. 3). As dextran is highly water soluble, >70% was released from 23 wt% gel, and 40% was released from 35 wt% gel, at day 12. We estimate that the release of a hydrophobic drug or protein drug from the gel will occur over one month.

Degradation of polymers also influences the release kinetics. A 23% aqueous solution of PEO–PLLA (M_r 5,000–1,000) diblock copolymers is a sol at 37°C. This diblock copolymer can be supposed to be the approximate average structure of the first

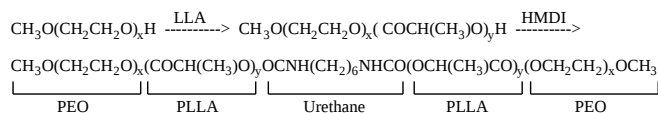


Figure 1 Synthetic scheme of block copolymers. Monomethoxy poly(ethylene oxide) (PEO; M_r 5,000) was dissolved in toluene, with residual water removed by azeotropic distillation, followed by polymerization of L-lactide (LLA) onto PEO using stannous octoate as a catalyst in toluene under reflux conditions. The coupling reaction of diblock copolymers was performed with hexamethylene diisocyanate (HMDI) in toluene at 60°C for 12 h, followed by reflux for 6 h. The triblock copolymers were purified by fractional precipitation of polymer from methylene chloride using diethyl ether.

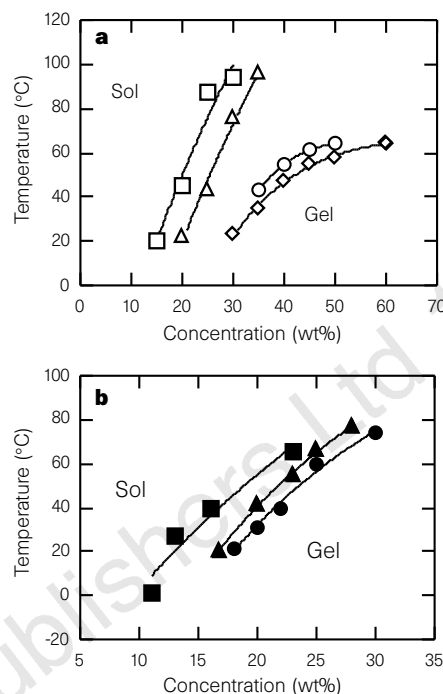


Figure 2 Gel-sol transition curves. **a**, PEO-PLLA diblock copolymers with M_r values as follows: diamonds, 5,000–720; circles, 5,000–1,000; triangles, 5,000–1,730; squares, 5,000–1,960. **b**, PEO-PLLA-PEO triblock copolymers with M_r values as follows: filled circles, 5,000–2,040–5,000; filled triangles, 5,000–3,000–5,000; filled squares, 5,000–5,000–5,000. The gel-sol transition temperature was determined as follows. The polymers were dissolved in distilled water at a given concentration in a tightly sealed vial which was stored at 4°C for 12 h; the vial was then incubated in a temperature-controlled bath for 20 min. The temperature of the bath was changed in steps of 2°C: conditions of gel (no flow) or sol (flow) were determined by inverting the vial vertically in the bath.

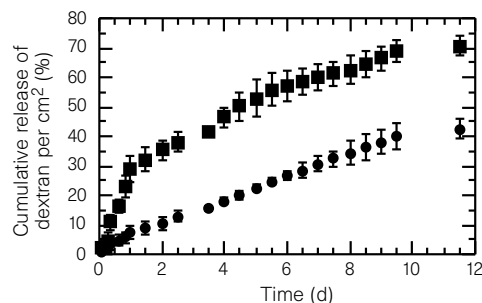


Figure 3 *In vitro* release profile of FITC-labelled dextran (M_r 20,000) from PEO-PLLA-PEO (M_r 5,000–2,040–5,000) triblock copolymer. FITC-labelled dextran (5.4 mg) was mixed with 0.5 ml of aqueous polymer solution (filled squares, 23 wt%; filled circles, 35 wt%). The mixture was injected into a 3.0-ml cuvette which was incubated in a 37°C water bath and 2.5 ml of distilled water was added. The area of the gel exposed to the water was 0.4 cm². With stirring, 1.0 ml of sample was taken at designated times and the same amount of distilled water was added. The amount of FITC-dextran released was calculated measurements of fluorescence intensity (excitation wavelength, 495 nm; emission wavelength, 515 nm) after dilution. Each data point represents an average $\pm$ standard deviation ($n = 3$).

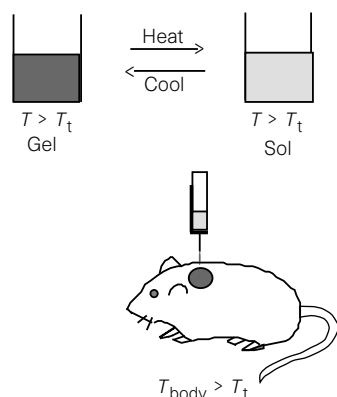


Figure 4 Injectable drug-delivery system. The sol phase containing bioactive agents forms a gel after subcutaneous injection. In the body, it can act as a depot for the controlled release of bioactive agents. T_t is the gel-sol transition temperature.

degradation product of PEO-PLLA-PEO (M_r 5,000–2,040–5,000) triblock copolymer if random scission of PLLA blocks is assumed to be the mechanism of degradation. We note that a 23% aqueous solution of PEO-PLLA-PEO (M_r 5,000–2,040–5,000) becomes a gel at 37 °C. This means one cleavage per triblock copolymer can lead to solubilization of the system.

Based on the thermosensitivity and biodegradability of these polymers, a solvent-free (no organic solvent) injectable system can be designed as a controlled drug carrier. Owing to the sharp phase transition, the sol becomes a gel after subcutaneous injection into the body. As an example, the sol state at 45 °C of PEO-PLLA-PEO (M_r 5,000–2,040–5,000) formed a gel upon subcutaneous injection into a rat (Fig. 4). As a gel in the body, it can act as a sustained-release matrix; this system has several advantages over common drug-delivery systems. First, the formulation is simple and requires no organic solvent. Second, the designed matrix can, if necessary, be stored at or below room temperature as a dry, solid form before administration. Last, there is no requirement for a surgical procedure for implantation. The administration of designed systems can be carried out by simple subcutaneous (or hypodermic-needle) injection after adding warm saline solution (45 °C). The subcutaneous injection of a sol at 45 °C makes causes neither significant pain nor tissue damage¹⁰ because the material is quenched to body temperature in a few seconds. Any pain could be avoided by using a local anaesthetic. The bioactive molecules entrapped in the polymer matrix will be released in the body at first by diffusion, and later by the combination of both diffusion and degradation mechanisms. PEO, PLLA and their degradation products are known to be biocompatible and pharmacologically inactive^{11,12}, so there is no need for removal of the implant. In contrast, poloxamers are non-biodegradable and enhance plasma cholesterol and triglycerol after intraperitoneal injection in rats¹³. The work that we report here strongly supports the use of copolymer gels for local delivery of bioactive agents, especially protein drugs, with a sustained action. □

Received 3 December 1996; accepted 1 July 1997.

1. Langer, R. New methods of drug delivery. *Science* **249**, 1527–1533 (1990).
2. Ishihara, K., Muramoto, N. & Shinohara, I. Controlled release of organic substances using polymer membrane with responsive function for amino compounds. *J. Appl. Polym. Sci.* **29**, 211–217 (1984).
3. Thomas, J. L., You, H. & Tirrel, D. A. Tuning the response of a pH-sensitive membrane switch. *J. Am. Chem. Soc.* **117**, 2949–2950 (1995).
4. Kwon, I. C., Bae, Y. H. & Kim, S. W. Electrically erodible polymer gel for controlled release of drugs. *Nature* **354**, 291–293 (1991).
5. Bae, Y. H., Okano, T., Hsu, R. & Kim, S. W. Thermo-sensitive polymers as on-off switches for drug release. *Makromol. Chem. Rapid Commun.* **8**, 481–485 (1987).

6. Chen, G. H. & Hoffman, A. S. Graft copolymers that exhibit temperature-induced phase transitions over a wide range of pH. *Nature* **373**, 49–52 (1995).
7. Malmström, M. & Lindman, B. Self-assembly in aqueous block copolymer solutions. *Macromolecules* **25**, 5440–5445 (1992).
8. Youxin, L. & Kissel, T. Synthesis and properties of biodegradable ABA triblock copolymers consisting of poly(L-lactic acid) or poly(L-lactic-co-glycolic acid) A-blocks attached to central poly(oxyethylene) B-blocks. *J. Controlled Release* **27**, 247–257 (1993).
9. Alexandridis, P., Holzwarth, J. F. & Hatton, T. A. Micellization of poly(ethylene oxide)-poly(propylene oxide)-poly(ethylene oxide) triblock copolymers in aqueous solution: Thermodynamics of copolymer association. *Macromolecules* **27**, 2414–2425 (1994).
10. Moritz, A. R. & Henrique, F. C. Jr Studies on thermal injury II. The relative importance of time and surface temperature in the causation of cutaneous burns. *Am. J. Pathol.* **23**, 695–720 (1947).
11. Ronneberger, B., Kao, W. J., Anderson, J. A. & Kissel, T. In-vivo biocompatibility study of ABA triblock copolymers consisting of poly(L-lactic-co-glycolic acid) A blocks attached to central poly(oxyethylene) B blocks. *J. Biomed. Mater. Res.* **30**, 31–40 (1996).
12. Zhu, K. J., Lin, X. & Yang, S. Preparation, characterization, and properties of polylactide (PLA)-poly(ethylene glycol) (PEG) copolymers: potential drug carriers. *J. Appl. Polym. Sci.* **39**, 1–9 (1990).
13. Wout, Z. G. M., Pec, E. A., Maggiore, R. H., Palcharia, P. & Johnston, T. Poloxamer 407-mediated changes in plasma cholesterol and triglycerides following intraperitoneal injection to rats. *J. Parenteral Sci. Technol.* **46**, 192–200 (1992).

Acknowledgements. We thank to G. G. Krueger for comments on dermatological issues and J. S. Kim for technical assistance. D. S. Lee was supported by the Ministry of Education of Korea. This work was supported by the University of Utah Research Foundation.

Correspondence and requests for materials should be addressed to S.W.K.

Influence of CO₂ emission rates on the stability of the thermohaline circulation

Thomas F. Stocker & Andreas Schmittner

Climate and Environmental Physics, Physics Institute, University of Bern, Switzerland

Present estimates of the future oceanic uptake of anthropogenic CO₂ and calculations of CO₂-emission scenarios¹ are based on the assumption that the natural carbon cycle is in steady state. But it is well known from palaeoclimate records^{2–5} and modelling studies^{6–9} that the climate system has more than one equilibrium state, and that perturbations can trigger transitions between them. Anticipated future changes in today's climate system due to human activities have the potential to weaken the thermohaline circulation of the North Atlantic Ocean^{10–12}, which would greatly modify estimates of future oceanic CO₂ uptake¹³. Here we use a simple coupled atmosphere-ocean climate model to show that the Atlantic thermohaline circulation is not only sensitive to the final atmospheric CO₂ concentration attained, but also depends on the rate of change of the CO₂ concentration in the atmosphere. A modelled increase to 750 parts per million by volume (p.p.m.v.) CO₂ within 100 years (corresponding approximately to a continuation of today's growth rate) leads to a permanent shut-down of the thermohaline circulation. If the final atmospheric concentration of 750 p.p.m.v. CO₂ is attained more slowly, the thermohaline circulation simply slows down. The reason for this rate-sensitive response of the climate system lies with the transfer of buoyancy in the form of heat and fresh water from the uppermost layers of the ocean into the deep waters below. This sensitivity of the simulated thermohaline circulation to the rate of change of atmospheric CO₂ concentration has potentially important implications for the choice of future CO₂-emission scenarios¹.

The burning of fossil fuels, cement production and changes in land use have led to an increase of the atmospheric concentration of CO₂ by almost 30% over the pre-industrial level of 280 p.p.m.v. (ref. 14). Carbon dioxide, and other gases such as CH₄ and N₂O, are the most important greenhouse gases after water vapour and decrease the outgoing longwave radiative flux at the top of the tropopause if their concentrations increase. Recent estimates from several inde-

pendent three-dimensional model simulations suggest that the global mean surface air temperature will increase by 2.1 °C to 4.6 °C for a doubling of atmospheric CO₂ concentrations¹. Examining an earlier suggestion¹⁵, Manabe and Stouffer¹¹ showed, using their model, that the Atlantic thermohaline circulation reduces in response to the warming and that a critical threshold value of CO₂ concentration lies between twice and four times the pre-industrial concentration, beyond which the Atlantic thermohaline circulation breaks down. Such permanent changes have a profound impact on the redistribution of heat in the Atlantic region of the Northern Hemisphere, in particular with respect to land–sea temperature contrast. Moreover, a decrease in ventilation and the heating of the sea surface can both reduce the CO₂-uptake of the ocean by as much as about 30% (refs 13, 16) assuming a steady-state natural carbon cycle.

Coupled three-dimensional atmosphere–ocean models are extremely expensive to run which precludes systematic sensitivity studies. In addition, climate drift and the coupling of the different components of these models still pose significant problems especially if future climate states are far from the control climate of these models. Here we use our simplified, three-basin zonally averaged ocean circulation model, coupled to a simple energy-balance model of the atmosphere^{17,18}. The model resolves the three ocean basins of the Pacific, Atlantic and Indian oceans which are connected by a circumpolar channel; the ocean dynamics are based on the principle of vorticity balance in a zonally averaged basin¹⁹. The atmospheric component is a simple energy-balance model¹⁸ which has been supplemented with an active hydrological cycle. This is done by solving the diagnostic equation of the meridional flux of latent heat, L , in the zonally averaged atmosphere²⁰;

$$\nabla \cdot \mathbf{L} = E - P \quad (1)$$

where $\nabla \cdot$ is the zonally averaged divergence operator in spherical coordinates, and E and P are evaporation and precipitation, respectively. For each latitude, evaporation is calculated as $E = E_1 + E_2 + E_3$, where E_i is the evaporation in ocean basin i

depending on basin sea surface temperature and the surface air temperature. The meridional flux of latent heat is taken to be proportional to the meridional moisture gradient, $L = -K\nabla q$, where K is a constant meridional eddy diffusivity depending on latitude and q is humidity. Following ref. 20 we assume a constant relative humidity, $r = 0.85$, and set $L = -Kr(\partial q_s/\partial T)\nabla T$, where q_s is the temperature-dependent saturation humidity and T is the surface air temperature. The zonally averaged precipitation P can then be determined from equation (1), but its distribution over the individual ocean basins i requires a closure assumption. Here the ratio P_i/P , where P_i is the precipitation over ocean basin i , is determined from the spin-up of the ocean model by restoring sea surface temperatures and salinities to observed zonal means. In the subsequent experiments, P_i/P is held constant at any latitude. The coupling between the ocean and the atmosphere component of the model involves heat and freshwater fluxes. Wind stress is held constant and associated stabilizing feedbacks^{21,22} are neglected here; also the seasonal cycle is not represented. In contrast to most three-dimensional coupled models, no flux correction is applied nor is it necessary to reach a stable steady state. This is important especially if large deviations from the original climate occur²³. The perturbation due to an increasing concentration of atmospheric CO₂ is incorporated into the energy balance using a constant climate sensitivity parameter²⁴ (see Fig. 1 legend).

In the present experiments the global mean equilibrium warming for $2 \times \text{CO}_2$ is 3.7 °C (exp. 560, Fig. 2) and 7.3 °C for $4 \times \text{CO}_2$. We have selected the climate sensitivity ΔF_{2x} such that the global mean surface air temperature difference for a CO₂ doubling, ΔT_{2x} , is consistent with ref. 12. This sensitivity is in the intermediate range of that reported in ref. 1. The increase of the atmospheric CO₂ concentration causes transient changes in our climate model that are very similar to those in three-dimensional coupled atmosphere–ocean models^{12,25}. The warmer surface air temperatures increase the saturation vapour pressure and lead to increasing freshwater fluxes in the North Atlantic and to warmer sea surface temperatures. Both

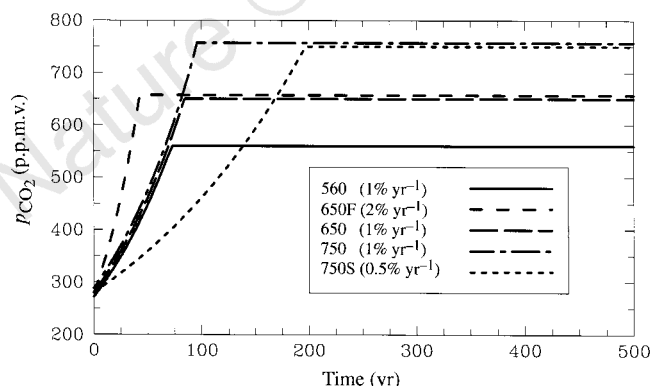


Figure 1 Prescribed evolution of atmospheric CO₂ (equivalent greenhouse gases) for five global warming experiments. The perturbation of the radiative flux in the energy balance model is a function of the CO₂ concentration: $\Delta F(t) = \Delta F_{2x} \log[p_{\text{CO}_2}(t)/280 \text{ p.p.m.v.}]/\log[2]$ according to ref. 24, where $\Delta F(t)$ and $\Delta F_{2x} = 7.3 \text{ W m}^{-2}$ are the time-dependent and $2 \times \text{CO}_2$ energy flux perturbations at the top of the atmosphere, respectively, and $p_{\text{CO}_2}(t)$ is the prescribed atmospheric concentration of CO₂ according to the specific scenario. The present scenarios agree with those of ref. 11. The standard rate of CO₂ increase is 1% yr⁻¹ compounded (approximately today's value); experiments with a fast rate of 2% yr⁻¹ (denoted F) and a slow rate of 0.5% yr⁻¹ (denoted S) are also given. The maximum p_{CO_2} values are 560 p.p.m.v. (exp. 560), 650 p.p.m.v. for experiments 650 and 650F, and 750 p.p.m.v. for experiments 750 and 750S. Once the maximum value is reached, $p_{\text{CO}_2}(t)$ is held constant. For better visibility, curves are slightly displaced where they would otherwise overlap.

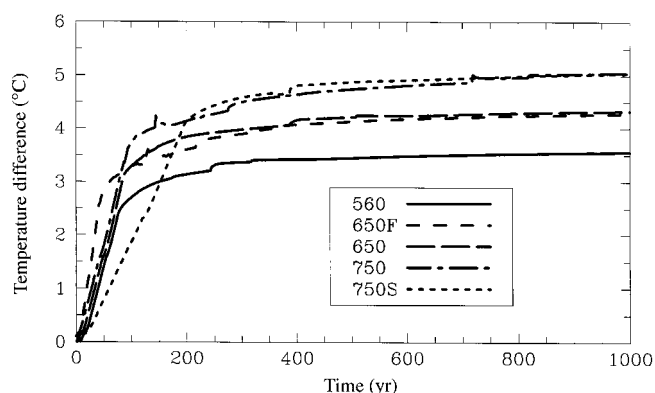


Figure 2 Simulated global mean surface air temperature changes for the five experiments given in Fig. 1. The equilibrium temperature difference is independent of the emission history for a given maximum CO₂ concentration. In these experiments $\Delta F_{2x} = 7.3 \text{ W m}^{-2}$. Note the considerable delay for final equilibration: it takes >1,000 years to distribute the excess heat in the deep ocean.

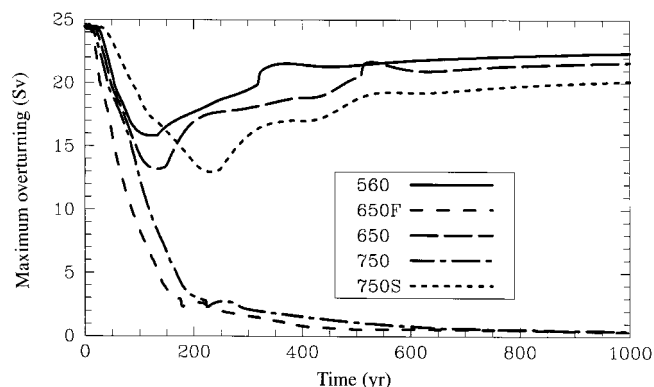


Figure 3 Evolution of the maximum meridional overturning of the North Atlantic in sverdrup ($1 \text{ Sv} = 10^6 \text{ m}^3 \text{ s}^{-1}$) for the five global warming experiments. In all cases, a reduction is obtained with an amplitude depending on the values of maximum atmospheric CO_2 and on the rate of CO_2 increase. The thermohaline circulation collapses permanently for a maximum concentration of 750 p.p.m.v. with an increase at a rate of $1\% \text{ yr}^{-1}$ (exp. 750). It recovers, however, and settles to a reduced value if the increase is slower ($0.5\% \text{ yr}^{-1}$, exp. 750S) or if the final CO_2 level is reduced to 650 p.p.m.v. (exp. 650). Similarly, for a fast increase (exp. 650F) at a rate of $2\% \text{ yr}^{-1}$ to only 650 p.p.m.v. the circulation collapses. All experiments have been integrated for 10,000 years and no further changes have been observed.

processes tend to increase the stratification of the water column and reduce the thermohaline circulation by a maximum of $\sim 34\%$ for a $2 \times \text{CO}_2$ experiment (exp. 560, Fig. 3) and a steady-state reduction of $\sim 10\%$. The sensitivity of the thermohaline circulation in this coupled model is therefore intermediate between that of ref. 25 (10%) and ref. 12 (50%). The timescales of the reduction of the circulation and of its subsequent recovery simulated in exp. 560 (Fig. 3) are in good agreement with those of the same experiment using a three-dimensional, coupled atmosphere–ocean general circulation model¹¹. This suggests that, in spite of the simplifications, the present model captures well the main large-scale processes that are relevant for these experiments.

With a CO_2 increase of $1\% \text{ yr}^{-1}$ (compounded, equivalent greenhouse gases) the thermohaline circulation collapses for maximum CO_2 concentrations of 700 p.p.m.v. and larger. This value depends on the climate sensitivity, $\Delta T_{2\times}$, and various model parameters. The circulation settles into a structurally different equilibrium state which is characterized by the complete absence of deep-water formation in the North Atlantic, much like today's Pacific Ocean. The physical mechanism for this transition is associated with a hysteresis behaviour observed when the Atlantic surface freshwater balance is changed^{9,26,27}. Further experiments showed that the increased stratification of the water column due to the warming only, that is, keeping $E - P$ constant in equation (1), is not sufficient to shut down deep water formation. This highlights the important role of changes of the hydrological cycle due to a warmer climate.

We found that in addition to the final CO_2 concentration, the stability of the ocean–atmosphere system is critically dependent on the rate of greenhouse-gas increase. For a slower increase (exp. 750S, Fig. 1) the Atlantic thermohaline circulation first reduces but then recovers, at a circulation weakened by $\sim 15\%$ (Fig. 3). Deep water is still being formed in the North Atlantic after CO_2 has reached its maximum value. As long as the Atlantic thermohaline circulation is not shut down, excess heat and fresh water, which tend to reduce the circulation, are still efficiently redistributed by advection, diffusion and especially convection. Therefore, the critical reduction of

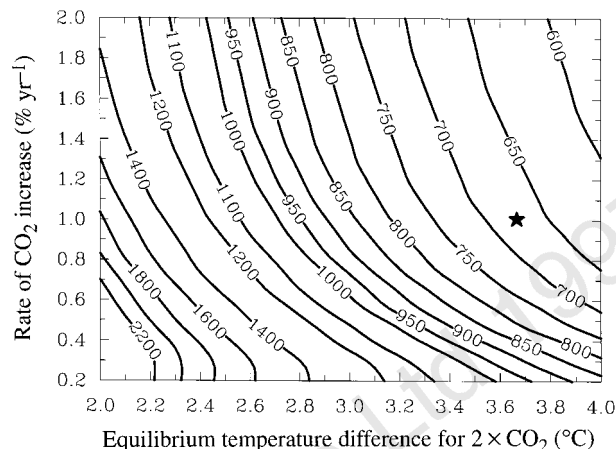


Figure 4 Threshold concentration of atmospheric CO_2 (contours in p.p.m.v.) beyond which the Atlantic thermohaline circulation breaks down permanently as a function of the equilibration surface air temperature difference for a doubling of CO_2 , $\Delta T_{2\times}$ (sensitivity of the climate model) and the rate of CO_2 increase in $\% \text{ yr}^{-1}$. The experiments in Figs 1–3 have been performed for $\Delta T_{2\times} = 3.7^\circ \text{C}$; for a CO_2 increase of $1\% \text{ yr}^{-1}$ the threshold lies between 650 and 700 p.p.m.v. in this model (denoted by a star). As for any climate model, the actual values depend on model parameters.

surface density, beyond which deep-water formation is inhibited, is not reached in this case. A slower warming thus tends to stabilize the system. But once the deep-water formation breaks down owing to a faster increase of atmospheric CO_2 and air temperature (Fig. 2), the subsequent process is self-sustaining: the absence of convection in the North Atlantic leads to an increased load of heat and fresh water in the upper layers of the ocean with a corresponding stronger stratification. A similar effect of the dependence of the stability of thermohaline circulation on the rate of perturbation changes was found in ocean-only models subject to localized freshwater injections^{26,28}.

There is still considerable uncertainty about the climate sensitivity of current climate models. We therefore investigated the threshold atmospheric CO_2 concentration beyond which formation of North Atlantic Deep Water is inhibited as a function of the equilibration temperature change for $2 \times \text{CO}_2$, $\Delta T_{2\times}$, and the rate of change of CO_2 in $\% \text{ yr}^{-1}$ (Fig. 4). This is the result of an extensive sensitivity study that is possible with this coupled climate model. For a given rate $\partial \Delta F / \partial t$ we made 1,000-yr experiments for increasing values of ΔF , the change in the radiation balance at the final CO_2 concentration, and determined the threshold value at which the thermohaline circulation shuts down permanently. For easier reference we then plot this result as contour lines of the threshold CO_2 concentration in a diagram where $p_{\text{CO}_2}^{-1}(\partial p_{\text{CO}_2} / \partial t)$ is plotted against $\Delta T_{2\times}$ (Fig. 4). For a typical climate sensitivity of $\Delta T_{2\times} = 3.7^\circ \text{C}$ and the present increase of CO_2 at $1\% \text{ yr}^{-1}$, the model yields a critical level of just below 700 p.p.m.v. For rates below the present value of $1\% \text{ yr}^{-1}$, the critical level is increasingly sensitive to the rate of CO_2 increase. A reduction of the rate of warming increases the critical level considerably. It is desirable to construct such a diagram with state-of-the-art three-dimensional coupled atmosphere–ocean models which contain more feedback processes and better spatial resolution than the present model. Furthermore, the implications of even a partial reduction of the Atlantic thermohaline circulation for biogeochemical cycles need to be investigated in greater detail.

The present results confirm an earlier study¹¹ which concluded that the atmosphere–ocean system contains critical levels of atmospheric CO₂ beyond which irreversible changes of the Atlantic thermohaline circulation are likely. Such changes would have also important direct and indirect consequences for the distribution of antropogenic carbon: a strong reduction of ventilation decreases the transport of excess carbon and heat into the deep ocean. The latter leads to an enhanced warming of the upper layers of the ocean which, in turn, decreases the solubility of CO₂ and results in a further increase of atmospheric CO₂. This feedback mechanism has the tendency to further destabilize the thermohaline circulation.

Our finding, that these critical levels also depend on the rate of CO₂ increase, is also relevant for decisions about the choice of a particular scenario of future greenhouse gas emissions. In the future, such decisions will not only need to consider the stabilization concentration or the economic cost of emission reduction to reach this concentration²⁹, but also take into account critical limits on the rate of greenhouse gas increase of the atmosphere. □

Received 21 April; accepted 30 June 1997.

- Houghton, J. T. et al. (eds) *Climate Change 1995: The Science of Climate Change* (Cambridge Univ. Press, 1996).
- Oeschger, H. et al. in *Climate Processes and Climate Sensitivity* (eds Hansen, J. E. & Takahashi, T.) 299–306 (Geophys. Monogr. 29, Am. Geophys. Un., Washington DC, 1984).
- Broecker, W. S., Petet, D. & Rind, D. Does the ocean–atmosphere system have more than one stable mode of operation? *Nature* **315**, 21–25 (1985).
- Broecker, W. S. & Denton, G. H. The role of ocean–atmosphere reorganizations in glacial cycles. *Geochim. Cosmochim. Acta* **53**, 2465–2501 (1989).
- Bond, G. et al. Correlations between climate records from North Atlantic sediments and Greenland ice. *Nature* **365**, 143–147 (1993).
- Bryan, F. High-latitude salinity effects and interhemispheric thermohaline circulations. *Nature* **323**, 301–304 (1986).
- Manabe, S. & Stouffer, R. J. Two stable equilibria of a coupled ocean–atmosphere model. *J. Clim.* **1**, 841–866 (1988).
- Maier-Reimer, E. & Mikolajewicz, U. *Experiments with an OGCM on the cause of the Younger Dryas 1–13* (Tech. Rep. 39, Max-Planck-Inst. für Meteorol., Hamburg, 1989).
- Stocker, T. F. & Wright, D. G. Rapid transitions of the ocean's deep circulation induced by changes in surface water fluxes. *Nature* **351**, 729–732 (1991).
- Manabe, S., Stouffer, R. J., Spelman, M. J. & Bryan, K. Transient responses of a coupled ocean–atmosphere model to gradual changes of atmospheric CO₂. Part I: Annual mean response. *J. Clim.* **4**, 785–818 (1991).
- Manabe, S. & Stouffer, R. J. Century-scale effects of increased atmospheric CO₂ on the ocean–atmosphere system. *Nature* **364**, 215–218 (1993).
- Manabe, S. & Stouffer, R. J. Multiple-century response of a coupled ocean–atmosphere model to an increase of atmospheric carbon dioxide. *J. Clim.* **7**, 5–23 (1994).
- Sarmiento, J. L. & Le Quéré, C. Oceanic carbon dioxide in a model of century-scale global warming. *Science* **274**, 1346–1350 (1996).
- Neftel, A., Oeschger, H., Staffelbach, T. & Stauffer, B. CO₂ record in the Byrd ice core 50,000–5,000 years BP. *Nature* **331**, 609–611 (1988).
- Broecker, W. S. Unpleasant surprises in the greenhouse? *Nature* **328**, 123–126 (1987).
- Maier-Reimer, E., Mikolajewicz, U. & Wingham, A. Future ocean uptake of CO₂: interaction between ocean circulation and biology. *Clim. Dyn.* **12**, 711–721 (1996).
- Wright, D. G. & Stocker, T. F. Sensitivities of a zonally averaged global ocean circulation model. *J. Geophys. Res.* **97**, 12707–12730 (1992).
- Stocker, T. F., Wright, D. G. & Mysak, L. A. A zonally averaged, coupled ocean–atmosphere model for paleoclimate studies. *J. Clim.* **5**, 773–797 (1992).
- Wright, D. G., Vreugdenhil, C. B. & Hughes, T. M. Vorticity dynamics and zonally averaged ocean circulation models. *J. Phys. Oceanogr.* **25**, 2141–2154 (1995).
- Chen, D., Gerdes, R. & Lohmann, G. A 1-D atmospheric energy balance model developed for ocean modelling. *Theor. Appl. Climatol.* **51**, 25–38 (1995).
- Weaver, A. J., Marotzke, J., Cummins, P. F. & Sarachik, E. S. Stability and variability of the thermohaline circulation. *J. Phys. Oceanogr.* **23**, 39–60 (1993).
- Schiller, A., Mikolajewicz, U. & Voss, R. The stability of the thermohaline circulation in a coupled ocean–atmosphere general circulation model. *Clim. Dyn.* **13**, 325–347 (1997).
- Egger, J. Flux correction: tests with a simple ocean–atmosphere model. *Clim. Dyn.* **13**, 285–292 (1997).
- Shine, K. P., Derwent, R. G., Wuebbles, D. J. & Morcrette, J.-J. in *Climate Change: The IPCC Scientific Assessment* (eds Houghton, J. Y., Jenkins, G. J. & Ephraums, J. J.) 41–68 (Cambridge Univ. Press, 1990).
- Johns, T. C. et al. The second Hadley Centre coupled ocean–atmosphere GCM: model description, spinup and validation. *Clim. Dyn.* **13**, 103–134 (1997).
- Mikolajewicz, U. & Maier-Reimer, E. Mixed boundary conditions in ocean general circulation models and their influence on the stability of the model's conveyor belt. *J. Geophys. Res.* **99**, 22633–22644 (1994).
- Rahmstorf, S. Bifurcations of the Atlantic thermohaline circulation in response to changes in the hydrological cycle. *Nature* **378**, 145–149 (1995).
- Rahmstorf, S., Marotzke, J. & Willebrand, J. in *The Warmwatersphere of the North Atlantic Ocean* (ed. Krauss, W.) 129–157 (Bornträger, Berlin, 1996).
- Wigley, T. M. L., Richels, R. & Edmonds, J. A. Economic and environmental choices in the stabilization of atmospheric CO₂ concentrations. *Nature* **379**, 240–243 (1996).

Acknowledgements. We thank N. Gruber, F. Joos, O. Marchal, H. Oeschger, S. Rahmstorf, A. Weaver and D. Wright for comments. This work was supported by the Swiss National Science Foundation.

Correspondence should be addressed to T.F.S. (e-mail: stocker@climate.unibe.ch).

Critical behaviour and the evolution of fault strength during earthquake cycles

Moritz Heimpel

Institut für Geophysik, Universität Göttingen, D-37075 Göttingen, Germany

The problem of how fault rheology and heterogeneity interact to produce the observed scaling of earthquakes (such as the power-law moment–frequency relationship) remains largely unsolved. Rock friction experiments have elucidated the properties of smooth faults^{1–3}, but seem insufficient to explain the observed complexity of real fault dynamics^{4,5}. The recognition of a connection between fault-related processes and critical phenomena in other physical systems, together with numerical models of repeated earthquakes, have resulted in significant progress in the theoretical interpretation of earthquake scaling^{4–14}. But fault rheology and heterogeneity have so far been treated separately. Here I attempt to unify the requirements of fault rheology and heterogeneity using numerical calculations of quantized slip in an elastic continuum. I show that cyclical fault strength evolves naturally by means of a statistical selection for high-strength fault patches (asperities), resulting in the accumulation and eventual failure of those asperities. The applicability of these results to real fault systems is supported by a recent analysis of time-dependent earthquake statistics¹⁵. These results imply that self-similarity and criticality on a fault emerge during an earthquake cycle, and suggest that the character of local seismicity can be useful in earthquake forecasting by revealing how advanced a fault is within its cycle.

During an earthquake the state of the fault–crust system changes rapidly. Stress is released and the fault strength drops. Between earthquakes, fault stress and strength recover through slip- and time-dependent healing. The physics of this cyclical process can be understood in terms of the evolution of a population of isolated contact surfaces^{16–18}. Such a population arises when two rough fault surfaces are pressed together with insufficient normal force to close apertures. An important parameter controlling this evolution is the minimum critical slip $L_{\min}$ required to change the strength of a population of contacts. Although fault surfaces are fractal and self-affine^{19–21}, it has been shown that two unmatched surfaces pressed together produce a maximum distance λ_c between contact surfaces (ref. 16), implying a minimum critical slip distance $L_{\min} \approx \lambda_c$. Estimates of λ_c (and thus $L_{\min}$) for natural fault surfaces typically range from 10^{-3} to 10^{-2} m (refs 16–18). For natural faults the presence of gouge also affects the scaling of $L_{\min}$; although not explicitly treated here, its effect is analogous to increasing fault surface roughness²².

In the numerical model presented here, λ_c (and thus also $L_{\min}$) is further interpreted to represent the minimum slip required to statistically decorrelate the local yield strength on a fault patch. This is clearly applicable if fault heterogeneity results from a diversity of loading and geometrical conditions on the scale of individual contacts. The model uses exact solutions for quantized slip on a fault surface embedded in a three-dimensional elastic medium. The model may be classified as inherently discrete (as defined in ref. 4) because, as discussed below, the yield strength of individual model cells can change abruptly. The main assumptions are as follows. (1) Inertia is neglected (quasi-static approximation). (2) Background stress increases uniformly over the fault plane. Time steps are chosen so that a quake initiates at a single numerical cell at every time step. (3) Slip is quantized; each of the $\sim 10^8$ discrete slips

in a simulation contributes an identical slip increment L and stress field σ_l to the fault domain. (4) Boundary conditions are periodic.

The model fault consists of 200×200 square cells. The shear stress at a cell centre is

$$\tau_k = \sum_l (\sigma_k)_l n_l + \sigma_0(t) \quad (1)$$

where $(\sigma_k)_l$ is the shear stress at cell location k due to the displacement discontinuity (that is, slip) Δu at cell l , n_l is the number of slip increments at cell l , and $\sigma_0(t)$ is the uniformly increasing shear stress. Periodic boundary conditions are imposed in the following way (see also ref. 4). The stress field due to each slip increment is

$$\sigma(i-i', j-j') = \sum_{p=-m}^{+m} \sum_{q=-m}^{+m} \sigma'(i-i'-p, j-j'-q) - \phi \delta(i-i'-p, j-j'-q) \quad (2)$$

where $\sigma'(i-i', j-j')$ is the shear stress at location (i, j) in the slip plane due to a single slip increment located at (i', j') in a three-dimensional full space, $p = q$ is the number of cells defining the width of the square numerical domain, ϕ is a dissipation parameter, and δ is the Kronecker delta function. For an infinite, periodically repeating stress field, $m = \infty$. Here, $m = 30$ is used as a good approximation.

There are several mechanisms by which faults dissipate energy, including viscous heating, distributed fracture and seismic radiation. However, to keep the interpretation of dissipation simple in the context of the quasi-static approximation, ϕ is set to release the amount of spatially averaged stress that would be lost out of a similar model fault patch with open boundaries, due to elastic stress transfer out of the model region. Thus

$$\phi = \left\langle \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} \sigma'(i-i', j-j') - \sum_{i=-p/2}^{p/2} \sum_{j=-q/2}^{q/2} \sigma'(i-i', j-j') \right\rangle \quad (3)$$

where the angle brackets indicate that the expression on the right-hand side of equation (3) is averaged over the model fault area. The boundary condition was chosen to best represent a sub-area on a fault of larger extent within the obvious limitations of computer modelling. The main disadvantage of periodic boundary conditions

is that wavelengths of deformation longer than the size of the numerical grid are neglected. An advantage is that, unlike the case of open boundary conditions, the stress drop for a single slip increment is independent of its position. As a result there is no time-averaged spatial bias for slip accumulation.

Fault heterogeneity is modelled using a yield-strength probability distribution

$$p(\sigma_y) = \frac{2\sigma_y}{\lambda^2} \left(1 + \frac{\sigma_y^2}{\lambda^2}\right)^{-2} \quad (4)$$

where σ_y is the yield strength and λ is a scale parameter such that $\mu = (\pi/2)\lambda$ is the mean. Equation (4) can be derived as a mixture of Weibull distributions and has been shown to be an appropriate model for fault heterogeneity²³. Note that σ_y may be interpreted either as the local yield strength or as a yield-strength fluctuation. This can be seen by writing $\sigma_y = \delta_0 + \sigma'_y$ where δ_0 is constant and σ'_y is the random variable. The choice of δ_0 simply sets the minimum strength and has no other effect on the results. In the numerical procedure λ (and hence μ) is constant and a standard algorithm using a random-number generator is used to output values of σ_y that are distributed according to equation (4). A model quake occurs when the strength of a cell obtains the yield strength, resulting in slip at that cell. Because slip changes the stress field on the numerical fault, one such slip can precipitate a large quake as a cascade of many (up to $\sim 10^7$) individual slips. For every quake each numerical cell that slips at least once is given a new strength from equation (4) after the initial slip. The new strength value is retained for subsequent slip during that quake. This *ad hoc* rule is meant to represent brittle failure of asperities, where the strength randomness represents the local heterogeneity associated with freshly fractured surfaces sliding past one another. Because the mean value μ of newly assigned strengths remains constant throughout the simulation, the probability of strengthening or weakening is proportional to the strength before failure.

The time evolution of domain-averaged fault properties is shown in Fig. 1. Here, a characteristic cycle is defined as the interval between events that rupture nearly the entire numerical domain. Although initial cell strengths are statistically uncorrelated in space and time, spatiotemporal correlations in model fault properties arise naturally from the interaction between the heterogeneity and the mechanics of the model system. Strong clusters of cells remain inactive longer before slipping; thus stronger asperities are selected

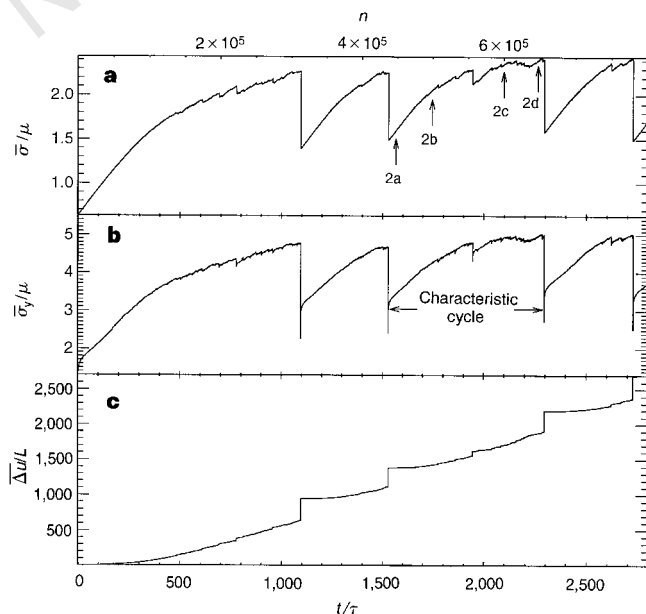


Figure 1 Domain-averaged time series of stress $\bar{\sigma}/\mu$ (a), yield strength $\bar{\sigma}_y/\mu$ (b) and slip $\bar{\Delta u}/L$, where $\bar{\Delta u}$ is the domain-averaged slip (c). A uniformly increasing background stress is imposed. The bottom horizontal axis is the dimensionless time t/τ , where τ is the time required to recover the stress drop due to a single slip increment over the entire fault domain. The top horizontal axis shows the step number n where each step corresponds to a model quake. Initially, yield strength is assigned by the distribution of equation (3), and stress is set at the mean strength μ . This implies that $\bar{\sigma}/\mu = 1$ and $\bar{\sigma}_y/\mu = 1$ at $t = 0$, where domain averages are indicated by an overbar. In the first time step, the average stress falls as all numerical cells with strengths $\sigma_y < \mu$ fail. The arrows marked by 2a, 2b, 2c and 2d indicate the times that correspond to the four slip images in Fig. 2. The calculation resulted in $n = 8 \times 10^5$ quakes with quake sizes ranging from $m = 1$ to $m \approx 10^7$ where m is the number of quantized slips per quake. Four full characteristic quake cycles occurred during this calculation. After an initial very large event at $n \approx 3 \times 10^5$, an asymptotic state obtains where increasing stress and strength are reset cyclically by three additional characteristic events that rupture almost the entire domain. A characteristic cycle begins with a very large quake, which precipitates a stress and strength drop, followed by relatively smooth stress recovery and strength increase (that is, healing) until a critical state is reached. Late in the cycle, large fluctuations in stress and strength correspond to larger model quakes that initiate local quake cycles on increasingly larger spatial scales, culminating in a very large event that initiates another characteristic cycle.

for and survive longer. As a result, during a quake cycle, a cell population evolves into one with an average yield strength that is significantly higher than the initial mean strength; the domain-averaged strength rises to a maximum of $\bar{\sigma}_y = 5\mu$.

The selection process during a characteristic cycle is shown in Fig. 2. Smooth, high-slip regions indicate low yield strength, whereas low-slip areas represent high-strength cell-clusters (asperities). The relative area occupied by asperities increases with time within the cycle.

Slip resulting from a moderately large quake sequence is shown in Fig. 3, illustrating several of the evolutionary phenomena described above. Figures 3a and b show quakes occurring at the latest stage of a cycle. The fore-shock (Fig. 3a) to main-shock (Fig. 3b) sequence shows localized slip on an increasing spatial scale that is representative of the model fault late in a characteristic cycle. The after-slip event (Fig. 3c) and the cumulative slip (Fig. 3d) show the transition to the early stage of a new cycle on the freshly slipped patch.

Figure 4 shows the time series of the domain-averaged stress-strength ratio $\bar{\sigma}/\bar{\sigma}_y$, clearly illustrating the stages of a model quake

cycle. To interpret the difference in size–frequency statistics between early (Fig. 4b) and later (Fig. 4c) stages, note that self-similar scaling in the rupture process is implied by a negative slope of $B = 2/3$ in the moment–frequency relation^{24,25}. Early in a quake cycle $B \approx 4/3$, indicating lack of self-similarity. Later in the cycle the fault reaches a critical stage where stress and strength fluctuations increase and B approaches $2/3$. These results support recent observations of time-variable microseismicity statistics at Parkfield, California, where the latest in an almost periodic (recurrence interval ~ 22 years) series of similar-sized large earthquakes (magnitude ~ 6) is overdue¹⁵. At Parkfield the B -value of microseismicity has decreased from 0.68 to 0.51 over about 4 years.

To summarize the model results, a characteristic quake cycle may be broken up into four stages. (1) Opportunism. The cycle begins with a flurry of relatively small aftershocks following the main-shock sequence of the previous cycle. The average yield strength increases rapidly with a decreasing slope as a fresh population of relatively weak asperities fail and are replaced by stronger ones. (2) Self organization. As the selection of relatively strong fault patches

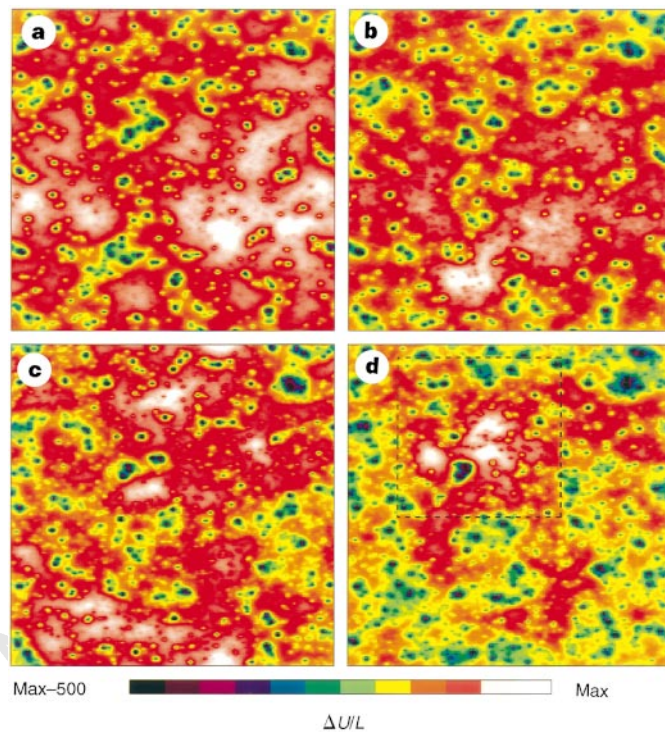


Figure 2 a–d, Images of relative amounts of total slip ΔU for the four moments in time marked in Fig. 1. Here, $\Delta U = \Delta u + \epsilon \Delta x$, where Δu is the slip discontinuity across the fault, ϵ is the shear strain, and Δx is the cell width. Each image shows the entire numerical domain and consists of 200×200 model cells. The colours for each image scale from the maximum ΔU (white) to 500L less than the maximum (black) for that image. The dashed square in **d** indicates the fault area shown in Fig. 3. **a**, The slip directly after a large event (that is, early in a characteristic cycle). The bright, relatively high-slip regions are 'smooth' as a result of extinction of asperities during the large event. Low-slip areas are indicative of high yield strength and clusters of low-slip cells may be referred to as 'asperities'. Each time a quake occurs, any numerical cell that slips is given a new and uncorrelated strength according to equation (4) (see explanation in text). During a cycle, slip preferentially occurs on weak fault cells so that strong cells and cell clusters have a greater mean lifetime. In this way asperities are selected for so that, late in a cycle, much of the model fault surface is composed of asperities (**d**). When asperities fail, the stress on the bounding fault surface increases in proportion to the integrated strength of the failed asperity. As a result of this proportionality, and the selection process, increasingly large fault patches fail during the course of a quake cycle.

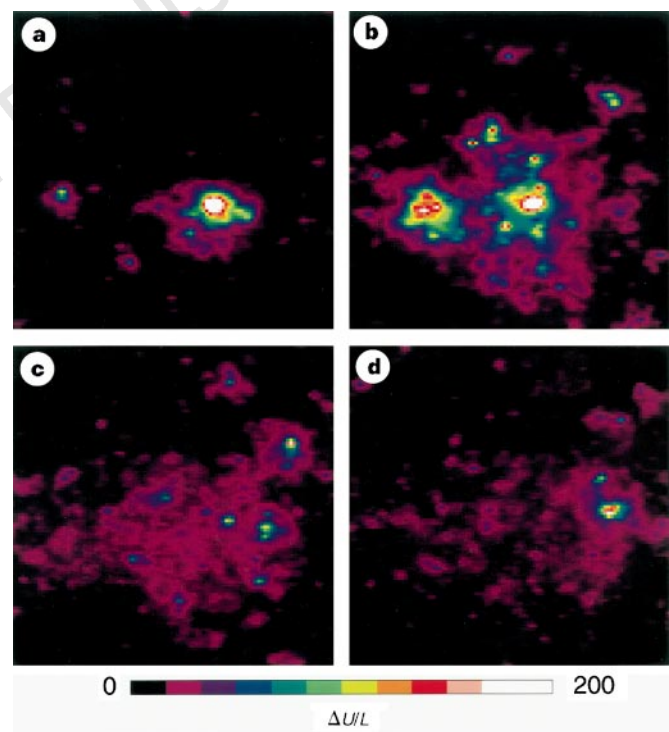


Figure 3 a–d Images of the total slip ΔU for a model quake sequence. The area for each image is indicated by the dashed square in Fig. 2d and consists of 100×100 model cells. **a**, A large model 'fore-shock', characterized by highly localized slip. This typically results when a cluster of high-yield-strength numerical cells (that is, an asperity) fails. **b**, The slip for the main shock. This event is precipitated by high stresses on the boundary of the fore-shock slip area. **c**, After-slip occurring on the main shock area, immediately following the main shock. This occurs because the main shock results in the replacement of asperities with cells of a mean strength distribution (given by equation (4)), so that the mean strength on that area is relatively low, and the strength distribution is relatively smooth compared to the surrounding fault area. As this area accumulates further slip, it will strengthen and roughen as asperities are selected for. **d**, Cumulative slip for 10 smaller 'after-shocks'. Model seismicity is shown to migrate away from the main shock area.

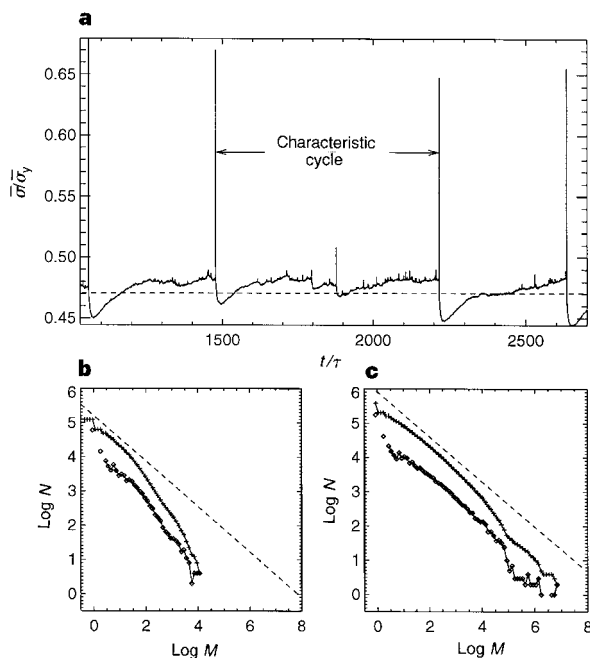


Figure 4 **a**, Domain-averaged ratio of stress to yield strength $\bar{\sigma}/\bar{\sigma}_y$ for $n = (3-8) \times 10^5$ for the last three characteristic cycles illustrated in Fig. 1. This plot clearly shows the stages within model quake cycles. Large events are indicated by sharp peaks in $\bar{\sigma}/\bar{\sigma}_y$. Also shown are the size-frequency statistics for the early stages of quake cycles ($\bar{\sigma}/\bar{\sigma}_y < 0.47$; **b**), and the later stages ($\bar{\sigma}/\bar{\sigma}_y > 0.47$; **c**). Immediately following large events, $\bar{\sigma}/\bar{\sigma}_y$ decreases rapidly as the weakest cells from a new population fail. The stress and strength continue to self-organize as fluctuations increase; the emergence of a critical stage is marked by a levelling-off or decrease of $\bar{\sigma}/\bar{\sigma}_y$ at a value above ~ 0.47 (indicated by the dashed line). Qualitatively, this occurs when high-strength asperities begin to fail. **b, c**, Plots of quake size (measured as $\log M$, where M is the absolute number of discrete slips per quake) versus the cumulative number of slips N (crosses) and N binned for every $1/10 \log M$ unit (diamonds). The size-frequency statistics for the subcritical state (defined as $\bar{\sigma}/\bar{\sigma}_y < 0.47$) and the critical state ($\bar{\sigma}/\bar{\sigma}_y > 0.47$) are shown in **b** and **c** respectively. The dashed lines in **b** and **c** show the negative slope $B = 2/3$ (corresponding to a Gutenberg-Richter 'b-value' equal to one). Comparing the slopes, it is apparent the Gutenberg-Richter scaling is more closely approximated during the critical than the subcritical state. This shows that self-similarity is violated in the early stages of a quake cycle and emerges during the transition to the critical state.

(asperities) continues, stress and strength increase. (3) Criticality. Here, large stress/strength fluctuations occur as stronger asperities fail, initiating four-stage quake cycles on increasing spatial and temporal scales within the characteristic cycle. (4) Extinction. The characteristic cycle ends with failure of major asperities and a main-shock sequence that spans nearly the entire domain. This is accompanied by a rapid stress and strength drop as the yield-strength distribution is reinitiated with the mean population strength. □

Received 15 January; accepted 21 July 1997.

- Dieterich, J. H. Modeling of rock friction I. Experimental results and constitutive equations. *J. Geophys. Res.* **84**, 2161–2168 (1979).
- Ruina, A. Slip instability and state variable friction laws. *J. Geophys. Res.* **89**, 10359–10370 (1983).
- Beeler, N. M., Tullis, T. E. & Weeks, J. D. The roles of time and displacement in the evolution effect in rock friction. *Geophys. Res. Lett.* **21**, 1987–1990 (1994).
- Rice, J. R. Spatio-temporal complexity of slip on a fault. *J. Geophys. Res.* **98**, 9885–9907 (1993).
- Rice, J. R. & Ben-Zion, Y. Slip complexity in earthquake fault models. *Proc. Natl Acad. Sci. USA* **93**, 3811–3818 (1996).
- Bak, P. & Tang, C. Earthquakes as a self-organized critical phenomenon. *J. Geophys. Res.* **94**, 15635–15637 (1989).
- Sornette, D. in *Spontaneous Formation of Space-Time Structures and Criticality* (eds Riste, T. & Sherrington, D.) 57–106 (Proc. NATO, ASI, Ser C, 349 Kluwer, Dordrecht, 1991).

- Xu, H. J., Bergersen, B. & Chen, K. Self-organized ruptures in an elastic medium: a possible model for earthquakes. *J. Phys. A* **25**, L1251–L1256 (1992).
- Rundle, J. B. & Klein, W. Scaling and critical phenomena in a cellular automaton slider-block for earthquakes. *J. Statist. Phys.* **72**, 405–413 (1993).
- Rundle, J. B., Klein, W. & Gross, S. Dynamics of a traveling density wave model for earthquakes. *Phys. Rev. Lett.* **76**, 4285–4289 (1996).
- Main, I. Statistical physics, seismogenesis, and seismic hazard. *Rev. Geophys.* **34**, 433–462 (1996).
- Burridge, R. & Knopoff, L. Model and theoretical seismicity. *Bull. Seismol. Soc. Am.* **57**, 341–371 (1967).
- Rydelek, P. A. & Sacks, S. Earthquake slip rise time and rupture propagation: Numerical results of the quantum earthquake model. *Bull. Seismol. Soc. Am.* **86**, 567–574 (1996).
- Ben-Zion, Y. Stress, slip, and earthquakes in models of complex single-fault systems incorporating brittle and creep deformations. *J. Geophys. Res.* **101**, 5677–5706 (1996).
- Nadeau, R. M. Characterization and application of microearthquake clusters to problems of scaling, fault zone dynamics, and seismic monitoring at Parkfield, California. Thesis, Univ. California, Berkeley, (1995).
- Scholz, C. H. The critical slip distance for seismic faulting. *Nature* **36**, 761–763 (1988).
- Scholz, C. H. *The Mechanics of Earthquakes and Faulting* (Cambridge Univ. Press, New York, 1990).
- Power, W. L., Tullis, T. E. & Weeks, J. D. Roughness and wear during brittle faulting. *J. Geophys. Res.* **93**, 15268–15287 (1988).
- Power, W. L. & Tullis, T. E. Euclidean and fractal models for the description of rock surface roughness. *J. Geophys. Res.* **96**, 415–424 (1991).
- Schmittbuhl, J., Gentier, S. & Roux, S. Field measurements of the roughness of fault surfaces. *Geophys. Res. Lett.* **20**, 639–641 (1993).
- Brown, S. R. & Scholz, C. H. Closure of rock joints. *J. Geophys. Res.* **91**, 4939–4948 (1985).
- Marone, C. & Kilgore, B. Scaling of the critical slip distance for seismic faulting with shear strain in fault zones. *Nature* **362**, 618–621 (1993).
- Heimpel, M. Earthquake size-frequency relations from an analytical stochastic rupture model. *J. Geophys. Res.* **101**, 22435–22448 (1996).
- Andrews, D. J. A stochastic fault model I. Static case. *J. Geophys. Res.* **85**, 3876–3877 (1980).
- Rundle, J. Derivation of the complete Gutenberg-Richter magnitude-frequency relation using the principle of scale invariance. *J. Geophys. Res.* **94**, 12337–12342 (1989).

Acknowledgements. I thank M. M. Wheelock for discussions. This work was supported in part by the Deutsche Forschungsgemeinschaft. Additional research was carried out during a visit to Johns Hopkins University, supported by the National Science Foundation.

Correspondence should be addressed to the author (e-mail: moritz@willi.uni-geophys.gwdg.de).

The Late Precambrian fossil *Kimberella* is a mollusc-like bilaterian organism

Mikhail A. Fedonkin* & Benjamin M. Waggoner†

* Paleontological Institute, Russian Academy of Sciences, ul. Profsoyuznaya 123, Moscow 117647, Russia

† Department of Integrative Biology, University of California, Berkeley, California 94720, USA

The fossil *Kimberella quadrata* was originally described from late Precambrian rocks of southern Australia¹. Reconstructed as a jellyfish², it was later assigned to the cubozoans ('box jellies'), and has been cited as a clear instance of an extant animal lineage present before the Cambrian^{3–7}. Until recently, *Kimberella* was known only from Australia, with the exception of some questionable north Indian specimens⁸. We now have over thirty-five specimens of this fossil from the Winter Coast of the White Sea in northern Russia. Our study of the new material does not support a cnidarian affinity. We reconstruct *Kimberella* as a bilaterally symmetrical, benthic animal with a non-mineralized, univalved shell, resembling a mollusc in many respects. This is important evidence for the existence of large triploblastic metazoans in the Precambrian and indicates that the origin of the higher groups of protostomes lies well back in the Precambrian.

The Winter Coast outcrops (Ust'-Pinega Formation) contain a rich 'Ediacara-type' biota^{9,10}. *Kimberella* occurs in several facies all through the section, but most specimens have come from the undersides of fine-grained sandstone gutter casts, in member 9 of the local stratigraphic section¹⁰. Channels in a clay substrate were colonized by a rich biota; when the channels were rapidly infilled with sand, the biota was preserved *in situ*¹¹. Associated with *Kimberella* in member 9 are large round 'medusoids', the distinctive fossils *Tribrachidium* and *Dickinsonia*, and some undescribed forms.

Small round fecal pellets, meandering trace fossils, and small pyritized algae also occur¹¹.

Kimberella is an oval-to-pear-shaped, bilaterally symmetrical fossil, with several zones arranged concentrically. Fossils range from 3 to 105 mm in length; the largest specimen is incomplete (Fig. 1c) and may have been 130–140 mm long in life. The size range is continuous; no discontinuities exist in size or structure that

would suggest sexual dimorphism or a size-structured population. Nor is there striking ontogenetic change over the observed size range; the smallest specimens are near-miniatures of the large ones. Films of framboidal pyrite on some specimens resulted from anaerobic bacterial activity after burial.

The distal portions of the fossils demonstrate a relatively inflexible outer margin, and an inner zone that is often distorted or folded

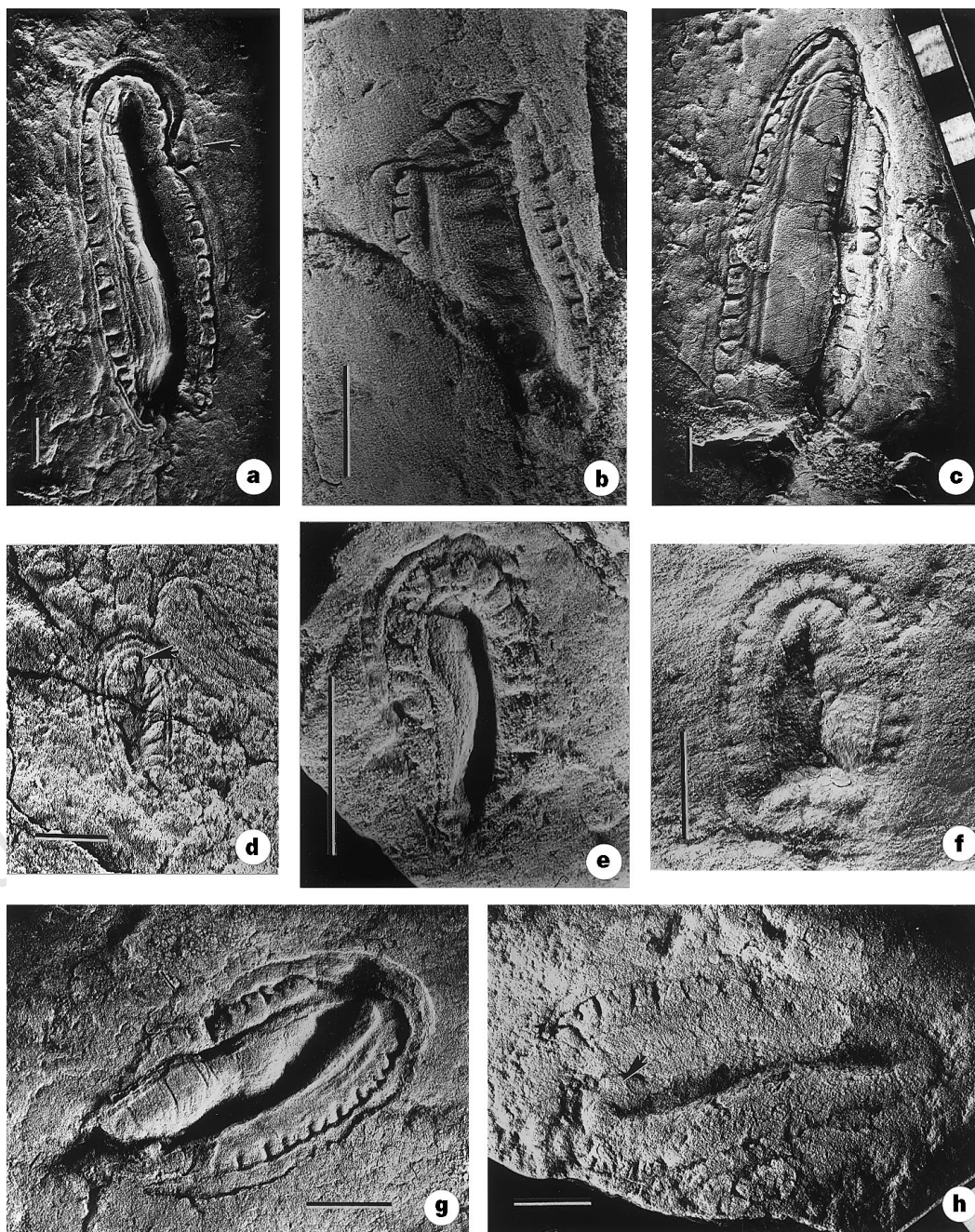


Figure 1 Specimens of *Kimberella* from Ust'-Pinega Formation, Winter Coast of White Sea, all coated with NH_4Cl . All specimens are curated at the Paleontological Institute (PIN), Russian Academy of Sciences, Moscow. All are preserved on undersides of sandstone channel casts in member 9 unless otherwise noted. All scale bars are 1 cm. **a**, Large, well preserved specimen, with deep central depression. Note folding of soft tissues away from shell edge, which has buckled (arrow) (PIN 3993/4003). **b**, Specimen partially folded over, demonstrating deformation of soft parts and bilateral symmetry of organism (PIN 3993/4026). **c**, Largest specimen known; note lateral displacement of shell, with crenellations extending out from under shell edge at right (PIN 3993/4004) (**d**). Specimen from

thinly interbedded argillite and siltstone, member 1 of local section; arrow indicates position of anterior bulge (PIN 3993/4001). **e**, One of two specimens showing series of pits (spicules?) along proximal ridge (PIN 3993/4036). **f**, Small but well preserved specimen showing primary features (PIN 3993/4033). **g**, Specimen showing rotation and displacement of shell from the main axis (PIN 3993/4027). **h**, Poorly preserved specimen from talus, probably from sandstones of member 10 of local section; note continuation of zones around posterior (pointed) end, as well as possible buccal mass indicated by arrow (PIN 3993/4009).

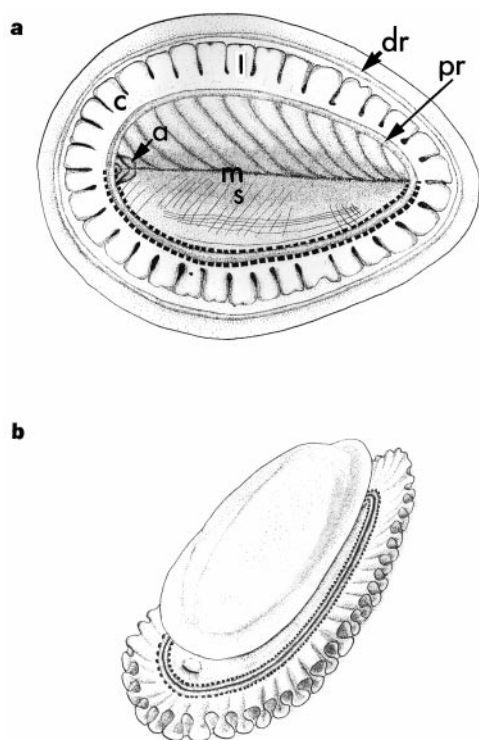


Figure 2 Reconstructions of *Kimberella*. **a**, Dorsal view. c, Crenellated zone; l, lobe; s, striae; dr, distal ridge; pr, proximal ridge; m, medial depression; a, anterior knoll. **b**, View of living organism, with folds of 'crenellated zone' extended beyond margin of shell; the folds would usually have been retracted under the shell at the time of burial.

over but never cracked (for example, see Fig. 1a, b, e). Within the inner zone is a row of crenellated structures, typically about 30 in number. These usually form a U-shaped band, but extend completely around the fossil in a few specimens. A thin linear structure, the proximal ridge, separates the crenellated zone from the interior; this ridge seems to be an overprint of a higher anatomical structure onto the imprint of the ventral surface of the organism. Rare specimens show a band of resistant point-like structures associated with the proximal ridge (Fig. 1e). The central part of the organism is preserved in negative relief. Its position and depth vary, depending on how much the organism was compacted and deformed; it may extend up to 20 mm into the host rock in large specimens. It may be compacted to be almost flat (Fig. 1f) or displaced with respect to the softer parts (Fig. 1c), but is rarely found bent laterally. Therefore, its original substance was comparatively firm¹². Numerous fine striae run laterally from the medial depression in many specimens; they are generally clearest in the most flattened specimens. At least three specimens show a small bulge near the broader end, on the midline of the body, inside the proximal ridge (Fig. 1d, h).

Kimberella was first formally described as a cubozoan medusa preserved on its side². The crenellated zones were interpreted as gonads or as muscle bands, and the organism was reconstructed as having tetradial symmetry. However, none of our specimens shows anything other than bilateral symmetry. If *Kimberella* were tetradially symmetrical, we would expect to find more than two crenellated zones in at least a few specimens. Even folded specimens show no more than two crenellated zones (see Fig. 1b for example). Nor was *Kimberella* triradially symmetrical¹³; the lateral zones are not repeated in the central zone. *Kimberella*, preserved in negative relief, was a 'resistant' organism; extant cubozoans are delicate and, if preserved at all, would form fossils in positive hyporelief, like most

'medusoids'¹². We cannot confirm the presence of gonads, tentacles and rhopalia⁵; the best-preserved specimens lack any trace of tentacles or rhopalia. The crenellated structures, formerly interpreted as gonads, formed a continuous band around the body, unlike the discrete gonads of cubozoans. Furthermore, *Kimberella* is common in member 9, which otherwise includes definitely benthic forms almost exclusively; there is no evidence of subaerial exposure or mass stranding at this level. Ediacaran fossils of planktonic organisms typically appear in different facies from benthic forms^{7,10}. We conclude that *Kimberella* was a benthic bilaterian, not a medusa.

The key to interpreting this fossil is the fact that the organism consisted of both firm and soft parts, and that compaction during decay 'overprinted' structures that in life were at different levels in the organism. In one specimen (Fig. 1g), both the central depression and the outermost ridge have been slightly rotated and moved forward together over the softer parts. Another (Fig. 1a) shows the softer 'crenellated structures' pulled away from the outer ridge, and another (Fig. 1c) shows them protruding beyond the outer ridge. Specimens like these show that the firmer parts formed a discrete anatomical unit, partially separable from the soft parts. We therefore interpret the central depression and outermost zone of *Kimberella* as parts of a thin oval shell, not mineralized but fairly stiff, at least in its central part. This shell lent the animal strength during burial and sediment compaction, at which time softer tissues were deformed and compressed by unlithified sediment entering the shell from below, and finally overprinted on the overlying sand.

The crenellated structures are often deformed and were clearly soft; we interpret them as lateral folds of the body, usually retracted under the shell but occasionally seen protruding (Fig. 1c). The frequent U-shape of the crenellated zone may be explained by the asymmetry of the antero-posterior outline of the shell (the highest point of the shell was closer to the anterior). The crenellations may have had a respiratory function. Although their number does not increase significantly with organism size, except in the largest specimens, the folds are small or even absent in small specimens, and longer and deeper in larger specimens. Assessing the respiratory function of the crenellations is difficult, because compaction makes it hard to estimate the volume of the organisms. However, our estimates of the total surface area of the crenellations scale isometrically with the product of body length and width. *Kimberella* may also have hosted microbial symbionts within the crenellations, as with the cerata or parapodia of symbiont-bearing gastropods¹⁴; symbiosis has often been proposed as a lifestyle for Ediacaran organisms^{13,15}. Some specimens show a broad flat region between the crenellated zones, which we interpret as a creeping foot. There is no sign of true segmentation of the body, but fine lines running from the midline to the foot in some specimens may be strands of dorsoventral musculature. The anterior bulge, identified in the Australian material as the 'stomach', may be the mouth and its associated muscles forming a buccal mass. There is no sign of a true radula, but radulae are very rarely preserved in any fossil molluscs.

Figure 2 shows our reconstruction of *Kimberella*. The basic architecture of the fossil, showing metamerism without observable segmentation, a shell and a broad flattened foot, is closely comparable to that of primitive molluscs. The arrangement and inferred function of the crenellations is like that of the ctenidia of chitons and the 'ventilatory flaps' of monoplacophorans, although as similar structures have evolved independently at least three times in the Mollusca¹⁶, the question of homology is not straightforward. The anatomy of *Kimberella* is also comparable to the soft anatomy of the Cambrian halkieriids, which appear to be relatives of molluscs and possibly other protostomes. *Halkieria* has a similar shape and division of the body into central and axial zones. It also has a zone of faint striae around the sole of its creeping foot, similar to the crenellated zone of *Kimberella* and plausibly having a respiratory function^{17,18}. It is even possible that the resistant pointlike structures

seen in rare specimens of *Kimberella* (Fig. 1e) are impressions left by sclerites or spicules like those of halkieriids and some molluscs, although no actual sclerites have yet been recovered. More material will be needed to test this hypothesis.

We cannot prove a molluscan affinity conclusively; some definitive molluscan synapomorphies, like the radula, cannot be seen, and others are open to other interpretations. However, no other extant metazoans resemble *Kimberella* as closely. Metamerism is consistent with a flatworm grade of organization, but the firm parts and complex anatomy of *Kimberella* tend to rule out a relationship to the Platyhelminthes. *Kimberella* somewhat resembles certain pelagic salps (phylum Urochordata), which have an oval shape, a tough thickened outer test, and serial muscle bands that could resemble the crenellations of *Kimberella*¹⁹. However, not only was *Kimberella* benthic, it lacks the atrial and branchial openings and the distinctive budding growth pattern of salps. The benthic habitat and absence of tentacles or obvious zooids rule out an affinity with siphonophores or chondrophorines. *Kimberella* superficially resembles ctenophores such as *Beroë*, but its resistant nature and lack of octamerous symmetry rule out a ctenophore affinity. Nor can we interpret *Kimberella* as a non-metazoan 'Vendobiontan'²⁰, protist²¹, or lichen²². It is always possible to stretch such hypotheses to cover almost any fossil imaginable, but we can find no definite features that would place *Kimberella* in any of these groups. *Kimberella* has nothing like the 'quilted pneu' morphology which has been used to argue against metazoan affinities for other Ediacaran fossils¹⁹, and its complex structural properties are not well explained by protist or lichen models.

We conclude that *Kimberella* is a bilaterian metazoan, more complex than a flatworm, more like a mollusc than like any other metazoan, and plausibly bearing molluscan synapomorphies such as a shell and a foot. This interpretation counters assertions that the Ediacara biota represents an extinct grade of non-metazoan life. It confirms hypotheses based on trace fossils that metazoan triploblastic lineages, including 'molluscan-grade bilaterians', began to diversify before the beginning of the Cambrian²³. It also suggests a pre-Ediacaran origin of major metazoan clades. □

Received 23 October 1996; accepted 27 May 1997.

- Glaessner, M. F. & Daily, B. The geology and Late Precambrian fauna of the Ediacara fossil reserve. *Res. S. Aust. Mus.* **13**, 369–401 (1959).
- Glaessner, M. F. & Wade, M. The Late Precambrian fossils from Ediacara, South Australia. *Palaeontology* **9**, 599–628 (1966).
- Wade, M. Hydrozoa and Scyphozoa and other medusoids from the Precambrian Ediacara fauna, South Australia. *Palaeontology* **15**, 197–225 (1972).
- Glaessner, M. F. *The Dawn of Animal Life* (Cambridge University Press, Cambridge, 1984).
- Jenkins, R. J. F. Interpreting the oldest fossil cnidarians. *Palaeontograph. Am.* **54**, 95–104 (1984).
- Jenkins, R. J. F. In *Origin and Early Evolution of the Metazoa* (eds Lipps, J. H. & Signor, P. W.) 131–176 (Plenum, New York, 1992).
- Gehling, J. G. The case for Ediacaran fossil roots to the metazoan tree. *Mem. Geol. Soc. India* **20**, 181–224 (1991).
- Shanker, R. & Mathur, V. K. Precambrian–Cambrian sequence in Krol belt and additional Ediacaran fossils. *Geophytology* **22**, 27–39 (1992).
- Fedonkin, M. A. in *The Vendian System* Vol. 1 (eds Sokolov, B. S. & Iwanowski, A. B.) 3–70 (Springer, Heidelberg, 1990).
- Fedonkin, M. A. Besskeletnaja fauna Vendi i ee mesto v evoljutsii metazoa. *Trudy Paleontologicheskogo Instituta* **227**, 1–174 (Nauka, Moscow, 1987).
- Grazhdankin, D. V. & Ivantsov, A. Yu. Reconstructions of biotopes of ancient Metazoa of the Late Vendian White Sea biota. *Paleontol. J.* **30**, 674–678 (1996).
- Wade, M. Preservation of soft-bodied animals in Precambrian sandstones at Ediacara, South Australia. *Lethaia* **1**, 238–267 (1968).
- McMenamin, M. A. S. & McMenamin, D. L. S. *The Emergence of Animals: the Cambrian Breakthrough* (Columbia University Press, New York, 1989).
- Smith, D. C. & Douglas, A. E. *The Biology of Symbiosis* (Arnold, London, 1987).
- Buss, L. W. & Seilacher, A. The Phylum Vendobionta: a sister group of the Eumetazoa? *Paleobiology* **20**, 1–4 (1994).
- Morton, J. E. in *The Mollusca 11: Form and Function* (eds Trueman, E. R. & Clarke, M. R.) 253–286 (Academic, San Diego, 1989).
- Conway Morris, S. & Peel, J. S. Articulated halkieriids from the Lower Cambrian of North Greenland. *Nature* **345**, 802–805 (1990).
- Conway Morris, S. & Peel, J. S. Articulated halkieriids from the Lower Cambrian of North Greenland and their role in early protostome evolution. *Phil. Trans. R. Soc. Lond. B* **347**, 305–358 (1995).
- Berrill, N. J. in *The Tunicata: with an Account of the British Species*. Figs 105, 106, 107 (Ray Society, London, 1950).
- Seilacher, A. Vendobionta and Psammocorallia: lost constructions of Precambrian evolution. *J. Geol. Soc. Lond.* **149**, 607–613 (1992).
- Zhuravlev, A. Yu. Were Ediacaran Vendobionta multicellulars? *N. Jb. Geol. Paläontol.* **190**, 299–314 (1993).

22. Retallack, G. J. Were the Ediacaran fossils lichens? *Paleobiology* **20**, 523–544 (1994).

23. Valentine, J. W. Late Precambrian bilaterians: grades and clades. *Proc. Natl Acad. Sci. USA* **91**, 6751–6757 (1994).

Acknowledgements. We thank the volunteers and colleagues who collected much of this material, in particular A. Yu. Ivantsov, D. V. Grazhdankin, J. Woland and A. G. Collins; Yu. I. Nikolaev, L. S. Zakharchenkova, the Paleontological Institute in Moscow, and the Arkhangel'skgeologiya office in Arkhangel'sk, for field logistical support; R. van Syoc (California Academy of Sciences) for access to specimens of extant invertebrates; D. R. Lindberg and J. G. Gehling for helpful comments; A. Mazin for photography and Y. A. Afanasiev for artwork.

Correspondence and requests for materials should be addressed to B.M.W. (e-mail: bmw@uclink2.berkeley.edu).

A neuronal population code for sound localization

Douglas C. Fitzpatrick, Ranjan Batra, Terrence R. Stanford* & Shigeyuki Kuwada

Department of Anatomy, University of Connecticut Health Center, Farmington, Connecticut 06030-3405, USA

The accuracy with which listeners can locate sounds is much greater than the spatial sensitivity of single neurons^{1–3}. The broad spatial tuning of auditory neurons indicates that a code based on the responses of ensembles of neurons, a population code, must be used to determine the position of a sound in space. Here we show that the tuning of neurons to the most potent localization cue, the interaural time difference in low-frequency signals (<~2 kHz; refs 4, 5), becomes sharper as the information ascends through the auditory system. We also show that this sharper tuning increases the efficiency of the population code, in the sense that fewer neurons are required to achieve a given acuity.

Sensitivity to interaural time differences (ITDs) first occurs in the main nuclei of the superior olivary complex (SOC). The information is then transmitted to the inferior colliculus, and from there to the auditory thalamus, which in turn projects to the auditory cortex. We examined ITD sensitivity at its point of creation (the SOC) and at two higher levels (inferior colliculus and auditory thalamus) through neural recordings in unanaesthetized rabbits.

Examples of the ITD tuning in the SOC and thalamus are shown in Fig. 1a. These 'composite' ITD curves represent the averaged responses to tones of different frequencies as a function of ITD⁶. In both nuclei, the curves are overlapping and broad in relation to the total range represented. Note, however, that the ITD tuning widths in the thalamus are much narrower than in the SOC.

The sharpening of ITD tuning as the information ascends the auditory system is also shown in Fig. 1b, c. The mean width of the composite curves showed a large decrease for the SOC to the inferior colliculus (730 to 430 μ s), and a smaller decrease from there to the thalamus (430 to 350 μ s) (Fig. 1b). The sharpening was not due to differences in the frequency content to which the samples of neurons in each structure responded, because it also occurred on a frequency-by-frequency basis (Fig. 1c). Much of the sharpening must therefore be due to neural processing. Possible mechanisms for the sharpening include inhibitory processes^{2,7} and convergence of temporally coincident activity⁸.

To determine how the sharpening of ITD tuning affects performance at each level, we used a model similar to that described in refs 9, 10. This model is similar to the 'position-variable' model of ITD sensitivity derived from behavioural data and responses from the auditory nerve¹¹. Such models begin by aligning all the neurons according to their preferred sensitivity, in this case the best ITD, to form a 'neural axis'. To the extent that the spatial field is not mapped

* Present address: Department of Neurobiology and Anatomy, Bowman Gray Medical School, Winston Salem, North Carolina 27157-1010, USA.

uniformly, the representation in the neural axis is distorted. In each nucleus, the distribution of best ITDs was peaked and biased towards ipsilateral (negative) delays, corresponding to sounds in contralateral space (Fig. 2a). The distorted neural representation that results is shown for the thalamus in Fig. 2b (top). The thalamic neurons are arrayed according to the best ITD, and the neural axis divided into 100 arbitrary units. The distortion in the map is evident, in that the first 40 neural axis units span 733 μ s of ITD, whereas the next 40 units span only 208 μ s; the remaining 20 units span 1,153 μ s.

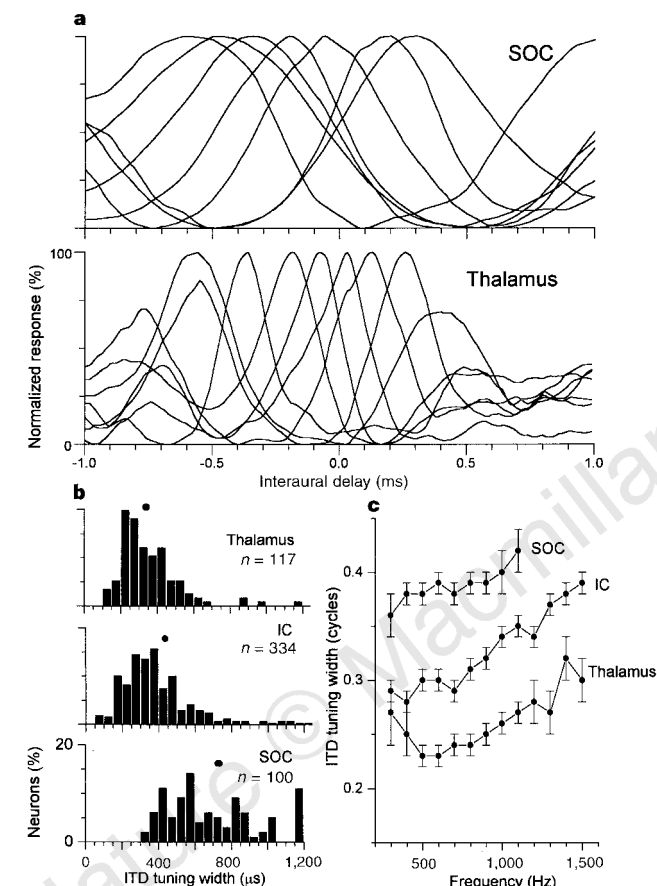


Figure 1 Sharpening of ITD tuning as the information ascends the auditory system. **a**, Examples of ITD tuning of neurons in the superior olivary complex (SOC) and auditory thalamus. Based on reconstructions from lesions made at the end of recording, the neurons in the SOC were primarily in the medial and lateral superior olives²⁹, whereas the thalamic neurons were located primarily in the posterior group of the thalamus and ventral and medial nuclei of the medial geniculate body³⁰. Each curve is the composite of responses to 'binaural-beat' stimuli (see Methods) of different frequencies that spanned the ITD-sensitive range for each neuron. The widths of the composite curves shown are near the median obtained for the SOC and auditory thalamus, respectively. The ITD tuning is much sharper in the thalamus than in the SOC. **b**, Distribution of composite curve widths measured at half-maximal response. The neurons at each level had a wide range of ITD tuning widths, but the mean (circles) shifted to progressively lower values as the information ascended. **c**, The widths of ITD curves to individual frequencies, measured at half-maximal response. Sharpening occurred on a frequency-by-frequency basis. Bars are standard errors based on the number of neurons sensitive to each frequency.

The performance of neurons at each level was assessed by using signal detection theory¹² to calculate the minimum detectable change in ITD that could be detected by different numbers of neurons. The response of each neuron to a range of target ITDs was obtained from their composite curves (Fig. 1a). In Fig. 2b (top), the responses of thalamic neurons ($n = 117$) to a target ITD of -100μ s (arrow) is represented by a series of vertical lines, one for each neuron, whose height is proportional to the maximal response of that neuron. By sampling from this distribution, we evaluated the ability of different numbers of neurons to discriminate changes in

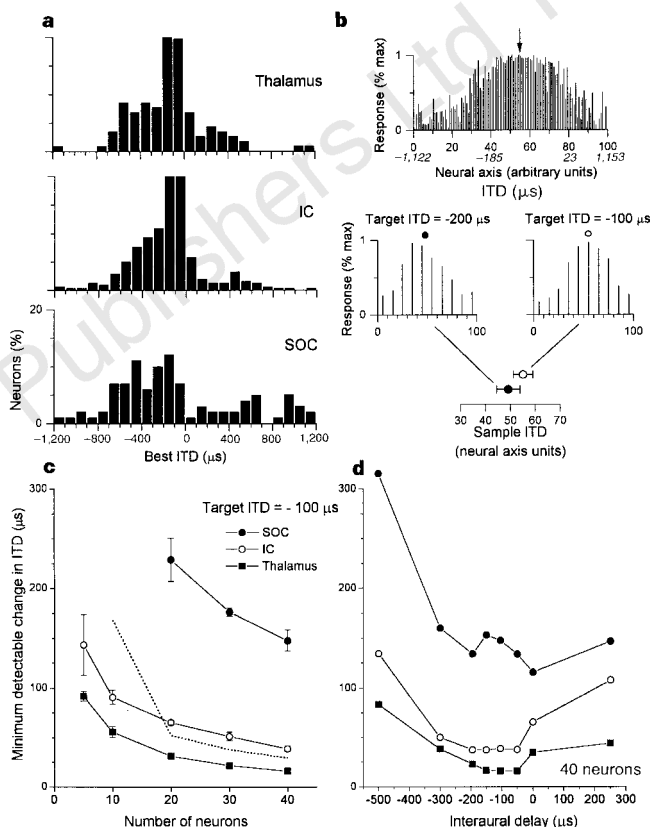


Figure 2 Construction and performance of a population model for ITD sensitivity. **a**, Distributions of best ITDs, or peaks of composite curves (Fig. 1a). These non-uniform distributions indicate that a portion of contralateral space is over-represented. **b**, Implementation of the model. The example shown is from the thalamus. Top, normalized responses of the neurons (vertical lines), arrayed according to their best ITD, to a target ITD of -100μ s (arrow). Middle, the responses to two target ITDs for a sample of 10 neurons. Neurons were randomly drawn, with replacement, from blocks evenly spaced along the ITD axis. A sample ITD to each target (circles) was determined according to $P = \sum_i (bITD_i^* r_i) / \sum_i r_i$, where $bITD_i^*$ was the best ITD of the i th neuron in neural ITD units, and r_i was the normalized response obtained from the composite curve. Bottom, means and standard deviations of sample ITDs for 100 samples of 10 neurons. A minimal detectable change in ITD was defined¹² as a d' of 1 (equivalent to about 75% correct responses), where $d' = \left| (P_1 - P_2) \right| / \left((s_1^2 + s_2^2) / 2 \right)^{1/2}$, with (P_1) and (P_2) equal to the mean sample ITD for each target and s_1 and s_2 the standard deviations. **c**, Minimum detectable change as a function of the number of neurons per sample. The points are the mean of 5 determinations of the minimum detectable ITD; the bars are the standard error. **d**, Minimum detectable change for 40 neurons to different target ITDs. At each level the acuity is highest in the regions of densest representation. The results are based on the results from one side only, because lesions in one hemisphere generally affect localization performance in contralateral space, as if responses from the two sides are not pooled.

ITD. Our method of evaluation is shown in Fig. 2b: the left middle panel represents the responses of 10 neurons drawn at random from one of 10 equally spaced blocks on the neural ITD axis to a target ITD of $-200 \mu\text{s}$ —the 'sample ITD' (circle) encoded by these neurons is given by the weighted average of the responses; the right middle panel shows the responses and sample ITD (circle) of the same neurons to a different target ITD ($-100 \mu\text{s}$). By repeating this process 100 times using different samples of neurons, the means and standard deviations of the sample ITDs were determined (Fig. 2b, bottom). These were compared for each sample size to determine a minimum detectable change in ITD (Fig. 2, legend). The aim of the model was to assess the efficiency of the population code for ITD representation at each level in terms of the number of neurons required to detect a given change in a stimulus ITD.

The results indicate that the effect of sharpening is to increase the efficiency of the population code. In Fig. 2c, we show the acuity achieved for different numbers of neurons at each level. The bars represent the standard error for five determinations of the minimum detectable ITD. For a target ITD of $-100 \mu\text{s}$, the acuity for all sample sizes was greatest in the thalamus, worst in the SOC, and intermediate in the inferior colliculus. For example, using 40 neurons a change of $147 \mu\text{s}$ could be detected in the SOC, $39 \mu\text{s}$ in the inferior colliculus, and $16 \mu\text{s}$ in the thalamus. Viewed from another perspective, to achieve a given acuity, fewer and fewer neurons were needed as the information ascended.

Further sharpening, however, does not necessarily result in a further increase in efficiency. When the thalamic neurons were artificially sharpened by an additional 50%, the performance decreased (Fig. 2c, dotted line). This effect is not unexpected, because in a noisy channel, sharper tuning beyond some optimal level will result in greater resolution of multiple stimuli but at the cost of increasing numbers of neurons to maintain the same degree of accuracy for all points in space^{13,14}. The decline in performance with artificial sharpening therefore suggests that the tuning of thalamic neurons is nearly optimal for maximum efficiency for the accurate representation of ITDs using a minimum of neurons. However, because the composite ITD curves are often complex (that is, they have sidebands; Fig. 1a), it was not clear whether the effect of that artificial sharpening was due to a decrease in peak width or to bringing the sidebands closer to the main peak. We therefore repeated the simulations using gaussian curves with ITD tuning and noise (d.c.) levels comparable to those of the recorded neurons, but without sidebands, and achieved a similar result. The auditory system therefore sharpens overly broad tuning from the SOC until it is appropriate for a maximally efficient population code. However, the final code still consists of elements that can be described as broadly tuned, in that they occupy a significant fraction of the total sensory field (on average $350 \mu\text{s}$ in a physiological range of $\sim 600 \mu\text{s}$, based on the rabbit's head width¹⁵).

An additional feature shown by the model is that performance at

each level was best over a range of ITDs in the centre of the distribution, and was markedly worse for more peripheral ITDs (Fig. 2d). In the region of ITDs near the edge of the physiological range ($-300 \mu\text{s}$ ITD), the performance was about half as good as in the centre. This feature has a behavioural correlate in humans, where sensitivity to changes in ITD declines by about a factor of two between the centre and the periphery¹⁶.

The increased performance at some ITDs at each level differs from a similar change seen in azimuthal visual spatial acuity. In the visual system, the sizes of receptive fields scale almost linearly with neural density. This relationship yields 'point images', defined as the amount of neural tissue representing any single point in the sensory field, which are nearly constant from centre to periphery. This organization is seen in visual cortex, where the definition is similar to that of the 'hypercolumn'¹⁷, and in the superior colliculus¹⁸. Higher acuity in the foveal representations arises because a smaller distance in visual space is required to move to a neighbouring module with a non-overlapping receptive field. By contrast, in the ITD system, neurons in the densest part of the distribution of best ITD are not markedly sharper than at more peripheral ITDs (Fig. 3a). This yields point images that are larger in the centre than at the periphery (Fig. 3b). This is equivalent to saying that receptive fields in the ITD system are not represented by equal-sized modules.

A consequence of point images that occupy a large fraction of available neural tissue is that such an organization is inherently poor at resolving multiple, simultaneous stimuli^{13,19}. Other systems based on similar broadly tuned neurons include the monkey motor cortex^{9,10}, the saccade system in the superior colliculus^{20,21}, and cricket cercal cells that measure wind direction²². In such cases population codes based on broadly tuned elements make considerable sense, because there can be only one arm movement, saccade or wind direction at a time. In contrast, the auditory system is faced at once with multiple, spatially segregated sources. This raises the question of how the locations of sources can be resolved neurally.

It could be that multiple sources are resolved along a dimension orthogonal to ITD, such as frequency. However, the average frequency tuning width (measured at half maximal response level) for thalamic neurons is 850 Hz, compared to the total frequency range for ITD sensitivity of $\sim 1,800 \text{ Hz}$ ($\sim 200 \text{ Hz}$ to 2 kHz). In fact, the change in frequency tuning as the information ascends is opposite to that of ITD tuning, in that it is narrowest in the SOC (average width, 560 Hz) and wider in the inferior colliculus (862 Hz) and thalamus. Thus neurons at higher levels of the ITD representation are broadly tuned to frequency as well as to ITDs, further indicating that the system is optimized to measure the ITD of one, or at most a few, sources at a time.

Human listeners exhibit a decline in behavioural performance in the presence of multiple stimuli with different ITDs. The early studies that showed very high acuity to ITD ($<20 \mu\text{s}$) used sequential rather than simultaneous stimuli¹. Direct comparisons

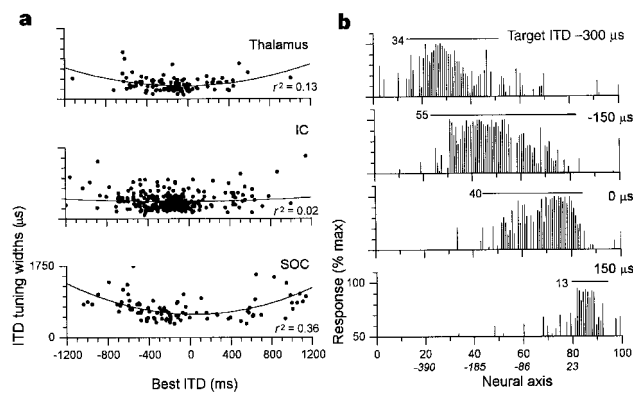


Figure 3 The point image representation of ITDs. **a**, Distribution of the widths of composite ITD curves, measured at half-maximal response, as a function of best ITD. There was little, if any, tendency for neurons in the region of greatest representation to have particularly sharp tuning, especially above the level of the SOC. **b**, Point images for different ITDs in the auditory thalamus. The point image (the portion of the neural axis activated by a particular ITD) can be appreciated graphically here as the extent of activation of neurons (vertical lines) above a threshold level (50% of maximal) to different target ITDs. The horizontal lines in each panel represent the calculated point image size (PI) in neural ITD units, where $PI = (RF_m)(MF)$, and RF_m is the mean 50% tuning width for a given target ITD (in μs), and MF is the magnification factor (in neural ITD units per μs of physical ITD) using a bin width of $100 \mu\text{s}$ for calculating RF_m and MF. Because the ITD tuning width does not decline with increased density of representation along the neural axis (**a**), the point images are larger in the centre of the distribution than in the periphery.

show a decline in acuity in the presence of multiple sources²³, as do earphone studies with stimuli containing 'interfering' ITDs^{24–26}. Our results suggest that the broadly tuned nature of the population code is the neural basis for this fundamental limitation. We suggest that to the extent we are aware of the locations of numerous, simultaneous sources, this 'auditory image' must be constructed on the basis of sequential determinations from different sources at different times. □

Methods

Animals ($n = 31$) were surgically prepared under ketamine/xylazine anaesthesia by attaching a bar to the skull with screws and dental cement. In a second surgery, a small hole was drilled in the skull overlying the cortex and filled with removable elastomer. Recordings were made without anaesthetic. The animals were restrained by securing the body in a close-fitting bag and clamping the head bar. The elastomer was removed, local anaesthetic was applied to the dura, and the electrode was directed towards an auditory centre. Daily recordings lasting ~2 hours were made over a period of up to several months. Usually only a single nucleus was studied. Recordings consisted of single neurons or clusters of a few neurons. Responses were obtained to 'binaural-beat' stimuli, which were pure tones (3.1 to 5.1 s long and repeated 2–4 times at a constant sound pressure level, usually near 70 dB) to each ear that differed in frequency by 1 Hz. This frequency difference created a cyclically varying ITD with a period of 1 s. Period histograms of the responses were created based on the 1-Hz binaural beat frequency; these yield the response as a function of interaural time difference. In our unanaesthetized preparation there was little if any difference in ITD tuning widths to static or dynamic (binaural beat) stimuli. Curves (Fig. 1a) were smoothed by counting spikes in 10 bins per cycle of the stimulating frequency. For all calculations used in later figures, 25 bins per cycle were used. The initial 100 ms was omitted to avoid onset effects. A full description of recording and stimulus conditions is provided elsewhere^{27,28}.

Received 18 February; accepted 4 June 1997.

- Mills, A. W. On the minimum audible angle. *J. Acoust. Soc. Am.* **65**, 991–1000 (1958).
- Kuwada, S., Batra, R. & Fitzpatrick, D. C. in *Binaural and Spatial Hearing in Real and Virtual Environments* (eds Gilkey, R. H. & Anderson, T. R.) 399–426 (Erlbaum, Mahwah, NJ, 1997).
- Brugge, J. F., Reale, R. A. & Hind, J. E. in *Binaural and Spatial Hearing in Real and Virtual Environments* (eds Gilkey, R. H. & Anderson, T. R.) 447–474 (Erlbaum, Mahwah, NJ, 1997).
- Blauert, J. *Spatial Hearing: The Psychophysics of Human Sound Localization* (MIT Press, Cambridge, MA, 1982).
- Wightman, F. L. & Kistler, D. J. The dominant role of low frequency interaural time differences in sound localization. *J. Acoust. Soc. Am.* **91**, 1648–1661 (1992).
- Yin, T. C. T. & Kuwada, S. Binaural interaction in low-frequency neurons in inferior colliculus of the cat. III. Effects of changing frequency. *J. Neurophysiol.* **50**, 1020–1042 (1983).
- Li, L. & Kelly, J. B. Inhibitory influence of the dorsal nucleus of the lateral lemniscus on binaural responses in the rat's inferior colliculus. *J. Neurosci.* **12**, 4530–4539 (1992).
- Stern, R. M. & Trahiotis, C. in *Auditory Physiology and Perception* (eds Cazals, Y., Horner, K. & Demany, J.) 547–554 (Pergamon, Oxford, 1992).
- Georgopoulos, A. P., Schwartz, A. B. & Kettner, R. E. Neuronal population coding of movement direction. *Science* **233**, 1416–1419 (1986).
- Georgopoulos, A. P., Kettner, R. E. & Schwartz, A. B. Primate motor cortex and free arm movements to visual targets in three-dimensional space. II. Coding of the direction of movement by a neuronal population. *J. Neurosci.* **8**, 2928–2937 (1988).
- Stern, R. M. & Colburn, H. S. Theory of binaural interaction based on auditory-nerve data IV. A model for subjective lateral position. *J. Acoust. Soc. Am.* **64**, 127–140 (1978).
- Green, D. M. & Swets, J. A. in *Signal Detection Theory and Psychophysics* 1–455 (Wiley, New York, 1966).
- Hinton, G. E., McClelland, J. L. & Rumelhart, D. E. in *Parallel Distributed Processing* Vol. I (eds Rumelhart, D. E. & McClelland, J. L.) 77–109 (MIT Press, Cambridge, MA, 1986).
- Salinas, E. & Abbot, L. F. Vector reconstruction from firing rates. *J. Comp. Neurosci.* **1**, 89–97 (1994).
- Heffner, H. & Masterton, B. Hearing in Glires: domestic rabbit, cotton rat, feral house mouse, and kangaroo rat. *J. Acoust. Soc. Am.* **68**, 1584–1599 (1980).
- Haftner, E. R. & DeMaio, J. Difference thresholds for interaural delay. *J. Acoust. Soc. Am.* **57**, 181–187 (1975).
- Hubel, D. H. & Wiesel, T. N. Receptive fields. Binocular interactions and functional architecture in the cat's visual cortex. *J. Physiol. (Lond.)* **160**, 106–154 (1962).
- Capuano, U. & McIlwain, J. T. Reciprocity of receptive field images and point images in the superior colliculus of the cat. *J. Comp. Neurosci.* **196**, 13–23 (1981).
- Rumelhart, D. E., Smolensky, P., Hinton, G. E. & McClelland, J. L. in *Parallel Distributed Processing* Vol. II (eds McClelland, J. L. & Rumelhart, D. E.) 77–109 (MIT Press, Cambridge, MA, 1986).
- Van Gisbergen, J. A. M., Van Opstal, A. J. & Tax, A. A. Collicular ensemble coding of saccades based on vector summation. *Neuroscience* **21**, 541–555 (1987).
- Lee, C., Rohrer, W. H. & Sparks, D. L. Population coding of saccadic eye movements by neurons in the superior colliculus. *Nature* **332**, 357–360 (1988).
- Theunissen, F. & Miller, J. P. Representation of sensory information in the cricket cercal sensory systems. II. Information theoretic calculation of system accuracy and optimal tuning curve widths of four primary interneurons. *J. Neurophysiol.* **66**, 1690–1703 (1991).
- Perrott, D. R. Concurrent minimum audible angle: a re-examination of the concept of auditory spatial acuity. *J. Acoust. Soc. Am.* **75**, 1201–1206 (1984).
- Trahiotis, C. & Bernstein, C. R. Detectability of interaural delays over select spectral regions: Effects of flanking noise. *J. Acoust. Soc. Am.* **87**, 810–913 (1990).

- Dye, R. H. The combination of interaural information across frequencies: Lateralization of the basis of interaural delay. *J. Acoust. Soc. Am.* **88**, 2159–2170 (1990).
- Heller, L. M. & Trahiotis, C. Extents of laterality and binaural interference effects. *J. Acoust. Soc. Am.* **99**, 3632–3637 (1996).
- Kuwada, S., Stanford, T. R. & Batra, R. Interaural phase sensitive units in the inferior colliculus of the unanesthetized rabbit. Effects of changing frequency. *J. Neurophysiol.* **57**, 1338–1360 (1987).
- Batra, R., Kuwada, S. & Stanford, T. R. Temporal coding of envelopes and their interaural delays in the inferior colliculus of the unanesthetized rabbit. *J. Neurophysiol.* **61**, 257–268 (1989).
- Batra, R., Kuwada, S. & Fitzpatrick, D. C. Sensitivity to interaural temporal disparities of low and high frequency neurons in the superior olivary complex. I. Heterogeneity of responses. *J. Neurophysiol.* (in the press).
- Stanford, T. R., Kuwada, S. & Batra, R. A comparison of the interaural time sensitivity of neurons in the inferior colliculus and thalamus of the unanesthetized rabbit. *J. Neurosci.* **12**, 3200–3216 (1992).

Acknowledgements. We thank C. Trahiotis, L. Bernstein and L. Abbott for discussion, L. M. Seman for technical assistance, and T. Ju and R. Manfredi for computer programming. This work was supported by a grant from the Center for Deafness and Communicative Disorders, NIH, USA.

Correspondence and requests for materials should be addressed to D.C.F. (e-mail: dcf@neuron.uchc.edu).

Loose-patch recordings of single quanta at individual hippocampal synapses

Lia Forti, Mario Bossi, Andrea Bergamaschi, Antonello Villa & Antonio Malgaroli

Department of Biological and Technological Research, DIBIT, Scientific Institute San Raffaele, and Cellular and Molecular Pharmacology Center, CNR, Department of Pharmacology, University of Milano, Via Olgettina 58, 20123 Milano, Italy

Synapses in the central nervous system are typically studied by recording electrical responses from the cell body of the post-synaptic cell. Because neurons are normally connected by multiple synaptic contacts, these postsynaptic responses reflect the combined activity of many thousands of synapses, and it remains unclear to what extent the properties of individual synapses can be deduced from the population response^{1–5}. We have therefore developed a method for recording the activity of individual hippocampal synapses. By capturing an isolated presynaptic bouton inside a loose-patch pipette and recording from the associated patch of postsynaptic membrane, we were able to detect miniature excitatory postsynaptic currents ('minis') arising from spontaneous vesicle exocytosis at a single synaptic site, and to compare these with minis recorded simultaneously from the cell body. The average peak conductance at a single synapse was about 900 pS, corresponding roughly to the opening of 90 AMPA-type glutamate-receptor channels. The variability in this conductance was about 30%, matching the value reported for the neuromuscular junction⁶. Given that our synapses displayed single postsynaptic densities (PSDs), this variability is larger than would be predicted from the random opening of receptor channels, suggesting that they are not saturated by the content of a single vesicle. Therefore the response to a quantum of neurotransmitter at these synapses is not limited by the number of available postsynaptic receptors.

Individual CA3–CA1 hippocampal boutons, visualized with the fluorescent dye FM1-43 (refs 7, 8), were enclosed inside patch pipettes and loose-seals⁹ were gently formed with the dendritic membrane underneath (Fig. 1a, b). In this configuration, quantal signals from synapses positioned outside the pipette are not visible above background noise (see Methods). Thus this approach provides a simple way of looking at currents from single synapses in isolation. For synaptic loose-seal recordings, we selected isolated chains of fluorescent boutons, belonging to putative single axons, above large proximal dendrites (~0–70 µm from the end of the soma) (Fig. 1a). The morphology of these contacts was analysed by serial electron microscopy (EM) of proximal dendrites. At these

locations, synapses were found to be discrete entities, with elliptically shaped varicosities (longitudinal diameter, $1.67 \pm 0.60 \mu\text{m}$; $n = 37$ synapses) filled with small synaptic vesicles (SSVs) (158.4 ± 91.7 SSVs; $n = 37$ serially reconstructed synapses), 12% of which were docked to the active region (18.6 ± 11.4 docked SSVs/synapse) (Fig. 1d). In 97% of cases ($n = 36/37$), presynaptic varicosities were apposed to individual postsynaptic densities (PSDs) having an average diameter of 441.6 ± 77.8 nm (Fig. 1c, d). Most of these synapses were composed of well differentiated spines and no glial layers were wrapped around the synaptic junctions (Fig. 1c). The average intersynaptic distance along individual axon collaterals was $4.42 \pm 1.91 \mu\text{m}$ ($n = 95$) (Fig. 1e), in agreement with measurements obtained by fluorescence confocal microscopy (5.12 ± 2.69 ; $n = 256$). Based on these morphological results, we conclude that an individual synapse can be contained inside a recording pipette with $\sim 2\text{-}\mu\text{m}$ tip, whereas the probability to accommodate two intact boutons inside is very low ($\sim 1\%$) (Fig. 1e).

Using this recording configuration (tetrodotoxin (TTX), AP5 (D-2-amino-5-phosphonovaleric acid) and picrotoxin always

present), we could detect AMPA-receptor-mediated miniature currents, always with inward polarity, in 18% of the synaptic sites investigated ($n = 22/125$) (Fig. 2a). At an individual site, minis exhibited variable amplitudes (mean peak amplitude = 53.6 ± 14.7 pA; $n = 11$) (Fig. 2a, b) but a remarkably similar time course (Fig. 2b). The time course of the average synaptic current was characterized by a fast rise ($t_{20-80\%} = 158 \pm 46 \mu\text{s}$; $n = 9$), followed by a slower decay, best described in most experiments by the sum of two exponential components (Fig. 2c, (time constants: $\tau_{\text{fast}} = 0.61 \pm 0.26$ ms; $\tau_{\text{slow}} = 2.55 \pm 1.19$ ms; peak amplitude ratio: $a_{\text{slow}}/a_{\text{fast}} = 0.33 \pm 0.23$; $n = 6$; $P < 0.025$). The presence of a biexponential decay time course could reflect either a complex glutamate kinetics in the synaptic cleft or different functional and/or molecular isoforms of AMPA channels at these synapses. To characterize these signals further, we carried out experiments in the presence of somatic voltage clamp. A somatic whole-cell (WC) recording was established before lowering the loose-patch pipette to enclose the selected fluorescent bouton and, as shown in Fig. 3a, when spontaneous events were detected, at a synaptic site, synchronous minis were

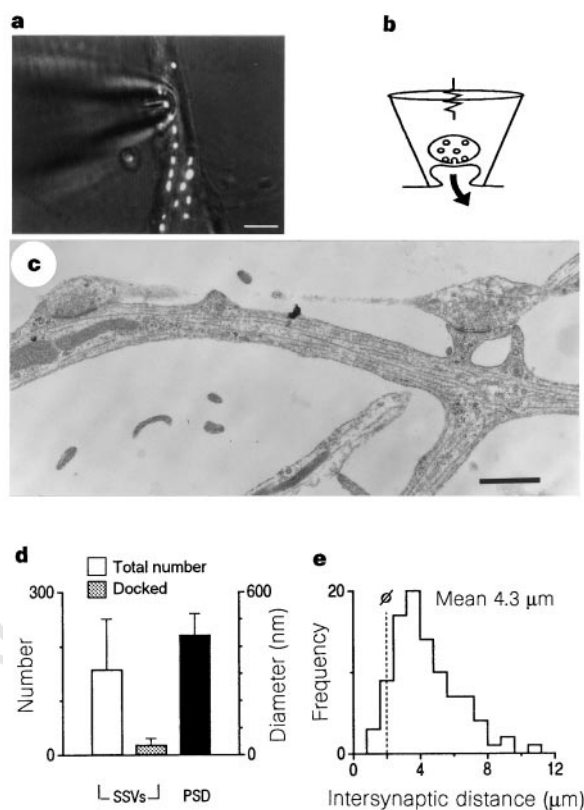


Figure 1 Loose-patch synaptic recordings. **a**, A fluorescence image of FM1-43 labelled synaptic boutons superimposed to the transmitted light view of a hippocampal neuron. A fluorescent synapse on a proximal dendrite is enclosed inside a patch pipette. Scale bar, $10 \mu\text{m}$. **b**, The recording configuration: the recording pipette forms a loose seal onto the postsynaptic membrane. **c**, Electron micrograph of an axon making two consecutive synapses along a proximal dendrite. Scale bar, $1 \mu\text{m}$. During loose-seal formation, the thin axon is presumably compressed between the pipette and the dendritic membrane. **d**, Numbers of total small synaptic vesicles (SSVs), docked SSVs/active region and longitudinal diameter of PSD from serially reconstructed synapses ($n = 37$ synapses; mean $\pm$ s.e.m.). **e**, Histogram of intersynaptic distances between centres of PSDs along individual axons as in **c** ($n = 95$). The compound probability for two intact boutons (longitudinal diameter, $1.67 \pm 0.60 \mu\text{m}$; $n = 37$) to fit inside the recording pipette (tip diameter (ϕ), $\sim 2 \mu\text{m}$) was $\sim 1\%$.

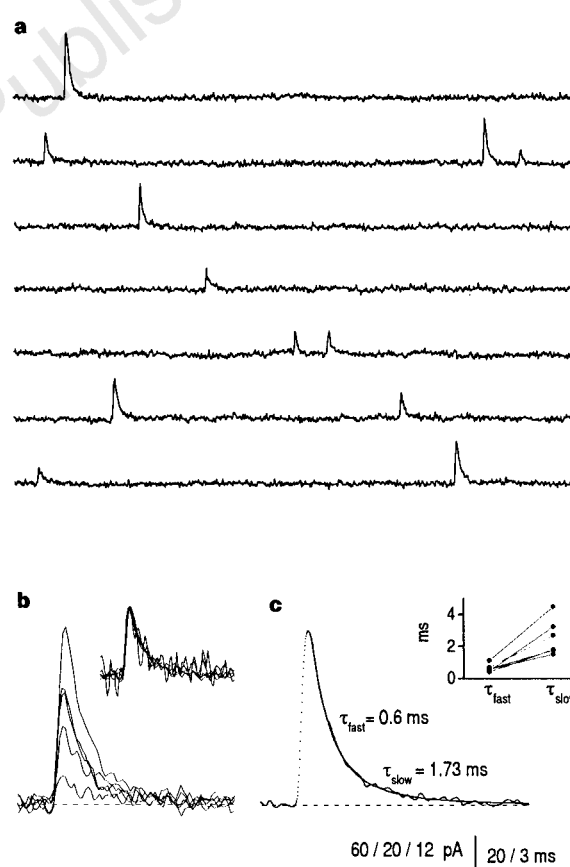


Figure 2 Miniature currents from single CA3-CA1 hippocampal synapses. **a**, Current traces from an exemplar single-synapse recording. Scale: 60 pA , 20 ms ($1 + R_{\text{pipette}}/R_{\text{seal}} = 1.33$; 2 kHz low-pass filtering; experiment LP008). **b**, Superimposition of 5 minis from the same experiment. Scale: 20 pA , 3 ms . Inset, the same events are shown after peak amplitude normalization to illustrate their almost identical time course. **c**, Average synaptic current from the same recording ($n = 14$ minis), with superimposed biexponential fit (solid line) to the decay phase ($a_{\text{slow}}/a_{\text{fast}} = 0.46$). Rise time ($t_{20-80\%}$) was $170 \mu\text{s}$. Scale: 12 pA , 3 ms . Inset, symbols plot the decay time constants (τ_{slow} and τ_{fast}) from 6/9 experiments where the decay timecourse was best described by the sum of two exponential components.

also present in the whole-cell trace ($n = 11/27$ experiments). In these experiments, focally recorded minis did not differ in amplitude and time course from those acquired in unclamped conditions. The amplitudes of single-synapse events detected simultaneously at the site of origin and at the soma were highly correlated, suggesting a linear propagation of minis along the dendritic cable (correlation coefficient, $r = 0.8 \pm 0.07$; $P < 10^{-4}$, $n = 4$ experiments) (Fig. 3b, inset). Across different experiments, a variable degree of amplitude attenuation (range, 0.29–0.96)

and slowing down of the time course was found (Fig. 3a, b), with no clear relation with either distance of the recorded sites from the soma or series resistance values ($R_s = 10$ – $20 \text{ M}\Omega$). Because, at individual synapses, quantal release events were consistently clustered in time (high- Ca^{2+} solution present in loose pipettes), in double recording experiments this produced correlated frequency fluctuations in quantal rate in both whole-cell and loose-patch traces (Fig. 3c). From these experiments, a direct estimate of the real synaptic conductance was obtained,

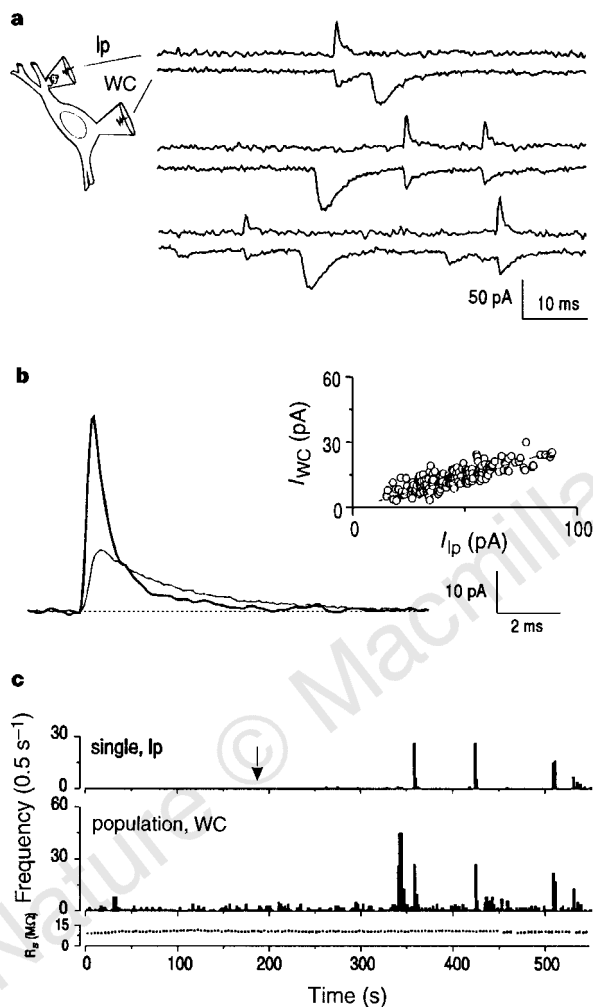


Figure 3 Simultaneous synaptic and somatic-wc recordings. **a**, Current traces from a loose-synaptic (lp; upper traces; $1 + R_{\text{pipette}}/R_{\text{seal}} = 1.53$) and somatic (WC; lower traces; holding potential, -55 mV) pair recording. Notice how minis originating from the synapse enclosed in the loose-patch pipette appear synchronously in both traces, whereas minis from other synapses are exclusively seen with the WC electrode (3 kHz low pass filtering; experiment LP064). **b**, Comparison of the average synaptic current from 130 paired minis recorded at the synaptic site (thick line; scaled by 1.53) or at the soma (thin line) (experiment LP064). Inset, correlation between amplitudes of single-synapse minis measured at the synaptic site (I_{lp}) and at the soma (I_{wc}) within the same experiment. The dotted line is a linear regression to all points ($R = 0.81$, slope = 0.29 , $P < 10^{-4}$, $R_s = 16 \text{ M}\Omega$). The linear relation between I_{lp} and I_{wc} suggests linear propagation of minis along the dendritic cable. **c**, Mini frequency time plots from a double-recording experiment. R_s for the WC recording is plotted below ($R = 11 \text{ M}\Omega$). Notice that the formation of the loose-seal (arrow) did not induce any visible perturbation in the WC recordings. Also, correlated frequency fluctuations are present in both plots, reflecting clustering of quanta at the recorded synapse (single, lp). Additional clusters appear in the somatic trace (population, WC) at $t \sim 30$ and $\sim 340 \text{ s}$ (experiment LP053).

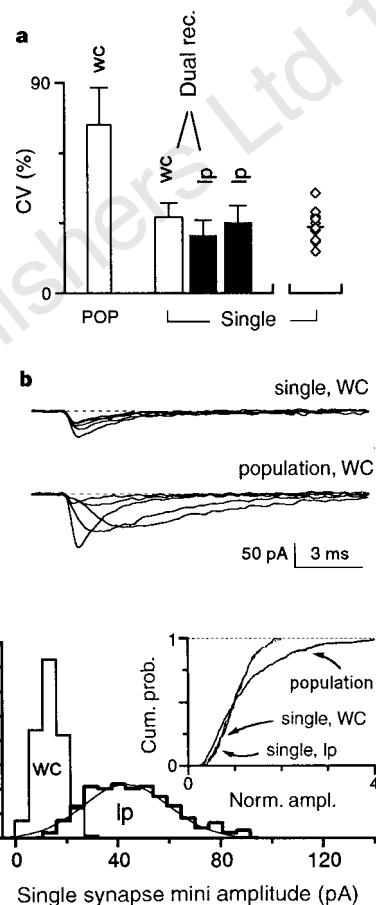


Figure 4 Quantal variability at single hippocampal synapses. **a**, Pooled values of amplitude CVs for population (Pop) and single-bouton minis (Single) (n , 7, 4, 4 and 7 experiments; mean $\pm$ s.d.). Minis were collected either focally (lp) or at the soma (WC) (Dual rec., simultaneous WC and synaptic recordings). CVs of population and single-bouton minis are significantly different ($P < 10^{-3}$, independent t -test). CV values for single-synapse minis, obtained under the various recording conditions, do not differ significantly from each other ($P > 0.05$, paired and independent t -tests). Symbols on the right plot all individual CV values ($n = 11$). The small horizontal line is their mean value. **b**, Single-synapse (top traces) and population minis (bottom traces) recorded at the soma. Note that single-synapse minis display a larger waveform similarity than do population events (experiment LP053). **c**, Amplitude histograms of 179 single-synapse minis measured focally (lp) (mean, 45.8 pA ; CV = 0.33 ; $\sigma_{\text{noise}} = 3.4 \text{ pA}$, 3 kHz low pass filtering) or after propagation to the somatic pipette (WC) (mean, 13.7 pA ; CV = 0.34 ; $\sigma_{\text{noise}} = 1.9 \text{ pA}$; 3 kHz filtering). The thin line is a gaussian function superimposed on the lp histogram (experiment LP064). Inset, cumulative plots of the same set of data superimposed to the amplitude distribution of population WC minis, recorded during the same experimental epoch (population) ($n = 685$; mean, 25.5 pA , CV = 0.68). Distributions were normalized to median values (Norm. Ampl.). Note the perfect overlap of the two single-synapse distributions (KS = 0.9) and the significant divergence from the distribution of population quanta (KS $< 10^{-3}$), which shows a pronounced positive skew (skew $_{wc=lp}$ = 1.94 and skew $_{wc=population}$ = 19.05 times the values expected for normally distributed data).

with a mean value of 886 ± 180 pS (range, 650–1,050 pS; holding potential (h.p.) = $-50/-60$ mV; $n = 4$). As the elementary conductance for AMPA channels in CA3–CA1 neurons is ~ 10 pS (ref. 10), this would translate into ~ 90 AMPA channels open at the peak of a mEPSC. This figure could be a slight underestimate, because in higher divalent concentrations a small reduction in AMPA conductance is expected¹¹.

When miniatures are collected from the somata of central neurons, a large quantal variability is consistently found^{3–5,12–14}, with broad and skewed amplitude distributions. This contrasts with the low unitary variability obtained from deconvolution of evoked distributions^{11,12,15–18} and many explanations for this discrepancy have been proposed^{1,2,4,11,19–21}. With synaptic loose-seal recordings, we found that the average degree of amplitude variability (coefficient of variation, CV = s.d./mean) was $28.4 \pm 7.2\%$ ($n = 11$), a value more than twofold lower than that found with standard somatic whole-cell recordings, in which the combined activity of many synapses was sampled ($72.4 \pm 15.5\%$; $n = 7$) (Fig. 4a). A large dilation of single-synapse CVs through synaptic fatigue processes, such as AMPA-receptor desensitization, is unlikely because in most experiments no correlation between mini amplitude and inter-event intervals was found ($P > 0.1$, $n = 8/11$; see Methods). In double-recording experiments (somatic whole cell and synaptic loose seal), we could directly compare properties of minis generated at individual synapses with those of events arising at the remaining synapses (population minis). In these experiments, single synapse minis displayed CVs significantly lower than those of population minis, both measured from somatic whole-cell traces ($CV_{wc-single} = 33 \pm 5.9\%$; $CV_{wc-population} = 53 \pm 11\%$; $P < 10^{-3}$, paired *t*-test; $n = 4$) (Fig. 4a). In addition, as shown in Fig. 4b, the time course of single-synapse minis was highly conserved, in contrast with the largely scattered waveforms of population minis, and this produced a significant difference in the respective waveform cross-correlation coefficients ($P < 0.05$; $n = 4$). When amplitude distributions for single-synapse minis were constructed, these displayed nearly gaussian shapes (Fig. 4c). Current propagation to the soma shifted amplitudes of single-synapse events towards lower values (Fig. 4c: WC versus Ip) but did not alter the shape of the underlying distributions; this was better visualized by superimposing the two normalized cumulative distributions (Fig. 4c, inset) ($KS > 0.9$, where *KS* is the Kolmogorov–Smirnov test on cumulative averages; $n = 4$ experiments). In the same set of experiments, a sharp divergence from the shape of amplitude distributions for population minis was revealed ($KS < 0.005$; as above), the latter exhibiting the characteristic pronounced skewness (Fig. 4c, inset) ($skew_{wc-single} = 2.1 \pm 1.7$ and $skew_{wc-population} = 15.2 \pm 4.8$ times the values expected for normally distributed data; $n = 4$ experiments). The small quantal variability observed at individual synapses cannot be attributed to the high- Ca^{2+} solution in loose-seal electrodes because in two synaptic recordings in which loose pipettes were filled with control Tyrode and some minis could be detected ($n = 2/37$), quantal variability matched values obtained in high- Ca^{2+} ($CV = 0.32$ and 0.43 ; $n = 21$ and 12 events). In addition, bath application of the same high- Ca^{2+} solution while recording minis from the cell soma (single-pipette whole-cell experiments) did not reduce but further increased the skewness of amplitude distributions ($skew_{wc-control} = 19.2 \pm 9.3$ and $skew_{wc-highCa2+} = 27.8 \pm 11.1$, as above), with a slight reduction in mini size ($12 \pm 21\%$), and in amplitude variability (17% ; $CV_{wc-control} = 81.5 \pm 10.6\%$; $CV_{wc-highCa2+} = 67.6 \pm 15.2\%$), which remained significantly higher than at single synapses ($P < 10^{-3}$). Taken together, these results indicate that the large variability of population minis (as detected by somatic recordings), rather than reflecting an intrinsic peculiarity of spontaneous transmission at one synapse, arises from the summed behaviour of many synaptic terminals. As in our system synapses on a given postsynaptic neuron display similar mean amplitudes⁴, then much of the observed

variability might originate from variable electronic filtering of synaptic currents injected at variable distances from the soma^{22,23}. Single-synapse CVs always in excess of the value expected from stochastic channel gating ($< 10\%$), in a condition where synapses display unitary PSDs, strongly suggests an incomplete occupancy of AMPA receptors after the release of one quantum^{24–27}.

In conclusion, our findings show that it is possible to record from a single central synapse and that its quantal properties can be accurately estimated. Our results show that quantal variability at individual glutamatergic synapses is far shorter than that seen in population recordings. Unimodal amplitude distributions and lack of receptor saturation suggest that minis, in this system, are likely to be unquantal events corresponding to the opening of ~ 90 AMPA channels. From this value, a rough estimate of 150–180 total available channels at an individual synaptic site can be obtained, assuming an open probability for AMPA channels of 0.5–0.6 at the peak of the synaptic event¹⁰. The high number of recordings from FM1–43 labelled synapses, where minis were not present (82%), besides the intrinsic technical difficulties could reflect lack of functional postsynaptic AMPA receptors, from enclosure inside loose pipettes of either GABAergic inhibitory boutons, or purely NMDA synapses^{28–30}. □

Methods

Synaptic loose-patch recordings. Postnatal CA3–CA1 hippocampal cultures were prepared as previously described¹⁴. Neurons (6–21 days old) were superfused ($22–25^{\circ}\text{C}$) with a solution containing (in mM): NaCl 119, KCl 5, $CaCl_2$ 2, $MgCl_2$ 2, HEPES 25, glucose 30; and picrotoxin, 100 μM (Sigma), APV 25–100 μM (Tocris), TTX 0.5 μM (Latexan), adjusted to 305 mOsm and pH 7.4 (control Tyrode solution). Synapses were labelled for 1–2 min with 10 μM FM1–43 (Molecular Probes) dissolved in modified Tyrode solutions (either 90 mM KCl or, in most experiments, 5 mM $CaCl_2$ /no added $MgCl_2$; 25 μM CNQX, Tocris, 0.5–1 mM kynurenate, Tocris). The recording pipette (tip diameter ~ 2 μm ; resistance, $R_{pipette}$, 1–2 M Ω), filled with a high- Ca^{2+} Tyrode solution ($CaCl_2$ 5 mM, no added $MgCl_2$; picrotoxin 100 μM , APV 25–100 μM , TTX 0.5 μM), was connected to a patch-clamp amplifier (Axopatch 200A; Axon) in voltage-clamp mode (VC) and held at the zero-current potential. Pipettes were lowered to enclose selected boutons and loose seals (seal resistance, R_{seal} , 2–11 M Ω) were obtained without suction. During recordings, 1-mV potential steps were applied every 2 s to monitor the stability of seal resistance. Experimental epochs accepted for amplitude analysis did not display seal resistance variation in excess of 5% of mean value and had less than 1 mV drift in ($V_{pipette} - V_{bath}$). With R_{seal} values of about two orders of magnitude larger than bath resistance, less than $\sim 1\%$ of ‘contaminating’ currents from synapses located outside the pipette would flow to ground through the electrode (below detection threshold, for signals up to a few hundred pA). The large frequency fluctuations in quantal rate in loose-seal recordings were induced by the high Ca^{2+} in the recording pipette as they were not seen in 37/37 loose-patch experiments with pipettes filled with control Tyrode (thus also excluding disruption of terminals with loose pipettes). All current traces were stored on a PCM-video recorder.

Somatic whole-cell recordings. Somatic voltage clamp recordings were obtained either in whole-cell or perforated patch configuration (amphotericin B from Sigma at 0.25 $\mu\text{g ml}^{-1}$) (EPC-7, List, or Axopatch 1D, Axon). Patch electrodes (2–5 M Ω) contained (mM): caesium gluconate 110, $MgCl_2$ 5, NaCl 10, EGTA or BAPTA 0.6–10, ATP 2, GTP 0.2, HEPES 49, adjusted to pH 7.2 and 290 mOsm. Input resistance (0.1–1.2 G Ω) and series resistance (R_s ; 5–20 M Ω) were continuously monitored for constancy and experiments were rejected if they varied by more than 15%. All minis were suppressed by application of CNQX (10 μM ; see also ref. 14).

Data analysis. Data were low-pass-filtered at 2–5 kHz and digitized at 10–70 kHz. Minis were detected semi-automatically using two threshold-crossing criteria on the current signal and on its first derivative (thresholds: 3 times background s.d. Computer simulations estimated the undetected events to be less than 3% for signal-to-noise ratios (*S/N*) > 4 and less than 20% for *S/N* between 3 and 4. In loose-patch recordings (background s.d. range: 2.8–5.4 pA) mini amplitude values used for mean amplitude estimate and histogram

display were corrected for current loss through the seal⁹ (correction factor $C = 1 + R_{\text{pipette}}/R_{\text{seal}}$; range 1.2–1.8). In all analysed regions, measured mini amplitude was stationary (no correlation between amplitude and time; $P > 0.1$). Mini amplitude CVs were computed from 18–179 (mean, 101 ± 54) events as $\sqrt{(\text{total variance of measured amplitudes} - \text{background variance})/\text{mean}}$; the estimated contribution of variance arising from current loss through the loose seal was negligible. No correlation between mini amplitudes and inter-mini intervals was found in 8/11 experiments ($P > 0.1$). In 3 experiments, a positive correlation was present and disappeared after removal of those few events with intervals below 15–20 ms, without any significant change in CV estimates. For averages of synaptic currents, events were aligned to $t_{1/2}$ of rise. The kinetics of focally recorded minis were slightly faster than for currents obtained with fast glutamate application on outside-out patches¹⁰, suggesting either a difference in AMPA receptors properties (extra-synaptic versus synaptic, slices versus cultures) or different glutamate time courses in the two conditions. Mini decays were fitted, using a nonlinear least-squares algorithm, with the sum of 1–2 exponential components (minimum number chosen by the likelihood ratio test). For analysis of waveform variability in double-recording experiments, couples of mini waveforms from the whole-cell trace, aligned to their starting point, were cross-correlated over 8 ms. Cross-correlation coefficients (ρ) for consecutive single synapse minis were compared with ρ values for consecutive population minis from the same trace using a Fisher z -transform and independent t -test ($P < 0.05$ considered as significant). Values throughout the text are means $\pm$ s.d. unless specified.

Electron and fluorescence microscopy. For standard EM, cell monolayers were washed with control Tyrode solution, fixed with 2% glutaraldehyde/4% paraformaldehyde in 0.1 M cacodylate buffer (45 min, 24 °C) and post-fixed with 2% OsO₄. After dehydration in ethanol and overnight infiltration (75% Epon-812, 25% ethanol), samples were embedded in Epon. Ultrathin serial sections (~ 60 nm; ≥ 4 sections per synapse; mean, 5.5) were doubly stained with uranyl acetate and lead citrate and examined with a Hitachi H-7000 microscope. Fluorescence measurements of intersynaptic distances were obtained after labelling with a polyclonal anti-synaptotagmin antibody using a confocal microscope (Biorad MRC1024). The fluorescence picture in Fig. 1a was obtained with an SIT camera (Hamamatsu C5985) attached to the recording microscope.

Received 23 January; accepted 6 June 1997.

- Jack, J., Larkman, A. U., Major, G. & Stratford, K. J. in *Molecular and Cellular Mechanisms of Neurotransmitter Release* (eds Stjärne, L. et al.) 275–299 (Raven, New York, 1994).
- Edwards, F. A. Anatomy and electrophysiology of fast central synapses lead to a structural model for long-term potentiation. *Physiol. Rev.* **75**, 759–787 (1995).
- Bekkers, J. M. & Stevens, C. F. NMDA and non-NMDA receptors are co-localized at individual excitatory synapses in cultured rat hippocampus. *Nature* **341**, 230–233 (1989).
- Liu, G. & Tsien, R. W. Properties of synaptic transmission at single hippocampal synaptic boutons. *Nature* **375**, 404–408 (1995).
- Lewis, C. A. & Faber, D. S. Inhibitory synaptic transmission in isolated patches of membrane from cultured rat spinal cord and medullary neurons. *J. Neurophysiol.* **76**, 461–470 (1996).
- Fatt, P. & Katz, B. Spontaneous subthreshold activity at motor nerve endings. *J. Physiol. (Lond.)* **117**, 109–128 (1952).
- Betz, W. J. & Bewick, G. S. Optical analysis of synaptic vesicle recycling at the frog neuromuscular junction. *Science* **255**, 200–203 (1992).
- Ryan, T. A. et al. The kinetics of synaptic vesicle recycling measured at single presynaptic boutons. *Neuron* **11**, 713–724 (1993).
- Stuehmer, W., Roberts, W. M. & Almers, W. in *Single Channel Recording* (eds Sakmann, B. & Neher, E.) 123–132 (Plenum, New York, 1983).
- Spruston, N., Jonas, P. & Sakmann, B. Dendritic glutamate receptors channels in rat hippocampal CA3 and CA1 pyramidal neurons. *J. Physiol. (Lond.)* **482**, 325–352 (1995).
- Jonas, P., Major, G. & Sakmann, B. Quantal components of unitary EPSCs at the mossy fiber synapse of CA3 pyramidal cells of rat hippocampus. *J. Physiol. (Lond.)* **472**, 615–663 (1993).
- Edwards, F. A., Konnerth, A. & Sakmann, B. Quantal analysis of inhibitory synaptic transmission in the dentate gyrus of rat hippocampal slices: a patch-clamp study. *J. Physiol. (Lond.)* **430**, 213–249 (1990).
- Silver, R. A., Traynelis, S. F. & Cull-Candy, S. G. Rapid-time-course miniature and evoked excitatory currents at cerebellar synapses *in situ*. *Nature* **355**, 163–166 (1992).
- Malgarioli, A. & Tsien, R. W. Glutamate-induced long term potentiation of the frequency of miniature synaptic currents in cultured hippocampal neurons. *Nature* **357**, 134–139 (1992).
- Larkman, A., Stratford, K. & Jack, J. Quantal analysis of excitatory synaptic action and depression in hippocampal slices. *Nature* **350**, 344–347 (1991).
- Stricker, A. C., Field, A. C. & Redman, S. J. Statistical analysis of amplitude fluctuations in EPSCs evoked in rat A1 pyramidal neurones *in vitro*. *J. Physiol. (Lond.)* **490**, 419–441 (1996).
- Kullmann, D. M. & Nicoll, R. A. Long-term potentiation is associated with increases in quantal content and quantal amplitude. *Nature* **357**, 240–244 (1992).
- Liao, D., Jones, A. & Malinow, R. Direct measurement of quantal changes underlying long-term potentiation in A1 hippocampus. *Neuron* **9**, 1089–1097 (1992).
- Bekkers, J. M., Richerson, G. B. & Stevens, C. F. Origin of variability in quantal size in cultured hippocampal neurons and hippocampal slices. *Proc. Natl Acad. Sci. USA* **87**, 5359–5362 (1990).

- Korn, H., Sur, C., Charpier, S., Legendre, P. & Faber, D. S. in *Molecular and Cellular Mechanisms of Neurotransmitter Release* (eds Stjärne, L. et al.) 301–322 (Raven, New York, 1994).
- Bauerfeind, R., Hüttner, W. B., Almers, W. & Augustine, G. J. Quantal neurotransmitter release from early endosomes? *Trends Cell Biol.* **4**, 155–156 (1994).
- Jack, J. J. B., Noble, D. & Tsien, R. W. *Electric Current Flow in Excitable Cells* (Oxford Univ. Press, Oxford, UK, 1983).
- Spruston, N., Jaffe, D. B., Williams, S. H. & Johnston, D. Voltage- and space-clamp errors associated with measurement of electrotonically remote synaptic events. *J. Neurophysiol.* **70**, 781–802 (1993).
- Faber, D. S., Young, W. S., Legendre, P. & Korn, H. Intrinsic quantal variability due to stochastic properties of receptor–transmitter interactions. *Science* **258**, 1494–1498 (1992).
- Hestrin, S. Activation and desensitization of glutamate-activated channels mediating fast excitatory synaptic currents in the visual cortex. *Neuron* **9**, 991–999 (1992).
- Silver, R. A., Cull-Candy, S. G. & Takahashi, T. Non-NMDA glutamate receptor occupancy and open probability at a rat cerebellar synapse with single and multiple release sites. *J. Physiol. (Lond.)* **494**, 231–250 (1996).
- Ferking, M. & Wilson, M. Saturation of postsynaptic receptors at central synapses? *Curr. Opin. Neurobiol.* **6**, 395–403 (1996).
- Liao, D., Hessler, N. A. & Malinow, R. Activation of postsynaptically silent synapses during pairing-induced LTP in CA1 region of hippocampal slice. *Nature* **375**, 400–404 (1995).
- Isaac, J. T. R., Nicoll, R. A. & Malenka, R. C. Evidence for silent synapses: implications for the expression of LTP. *Neuron* **15**, 427–434 (1995).
- Durang, G. M., Kovalchuk, Y. & Konnerth, A. Long-term potentiation and functional synapse induction in developing hippocampus. *Nature* **381**, 71–75 (1996).

Acknowledgements. We thank D. DiFrancesco, D. S. Faber, A. Marty, J. Meldolesi and R. W. Tsien for comments on the manuscript, and A. Abenavoli, L. Anghinoni, M. Caimi, P. Cerutti, E. Naldi and J. Noel for help and discussion. L.F. was the recipient of a postdoctoral fellowship from the University of Milano. This research was supported by grants from EC Biomed-2, Telethon and PNR/Neurobiology (MURST-Italy) to A.M.

Correspondence and requests for materials should be addressed to A.M. (e-mail: mailgara@dibit.hsr.it).

Microglial activation by Alzheimer amyloid precursor protein and modulation by apolipoprotein E

Steven W. Barger*†‡ & Ashley D. Harmon†

* Donald W. Reynolds Department of Geriatrics, and † Department of Anatomy, Cell Biology, and Neurobiology, University of Arkansas for Medical Sciences, Little Rock, Arkansas 72205, USA

Geriatric Research Education and Clinical Center, John L. McClellan Memorial Veterans Affairs Medical Center, Little Rock, Arkansas 72205, USA

A role for β -amyloid precursor protein (β -APP) in the development of Alzheimer's disease has been indicated by genetics¹, and many conditions in which β -APP is raised have been associated with an increased risk of Alzheimer's disease or an Alzheimer's-like pathology^{2–4}. Inflammatory events may also contribute to Alzheimer's disease⁵. Here we investigate whether a secreted derivative of β -APP (sAPP- α) can induce inflammatory reactions in microglia, which are brain cells of monocytic lineage. We found that treatment with sAPP- α increased markers of activation in microglia and enhanced their production of neurotoxins. The ability of sAPP- α to activate microglia was blocked by prior incubation of the protein with apolipoprotein E3 but not apolipoprotein E4, a variant associated with an increased risk for Alzheimer's⁶. A product of amyloidogenic β -APP processing (sAPP- β) also activated microglia. Because sAPP- β is deficient in the neuroprotective activity shown by sAPP- α , our results indicate that increased amyloidogenic processing could adversely affect the balance of sAPP activities that determine neuronal viability.

To determine the impact of sAPP on inflammatory reactions in microglia, we treated the N9 microglial cell line⁷ with sAPP and measured the activity of the transcription factor NF- κ B by electrophoretic mobility-shift assay (EMSA) (Fig. 1). These cells responded to sAPP with activation of NF- κ B within 90 min. Induction of a κ B-binding transcription factor by sAPP in primary neurons is dependent upon an increase in the concentration of cyclic GMP⁸ and involves a transcription factor distinct from NF- κ B (unpublished

results). However, the addition of an inhibitor of cGMP-dependent protein kinase did not block the activation of NF- κ B by sAPP in N9 cells, and a cell-permeant cGMP analogue did not mimic the effects of sAPP in these cells. These results indicate that the activation of NF- κ B by sAPP in microglial cells occurs through a cGMP-independent mechanism and could thus involve signalling events evoked by regions of sAPP distinct from the carboxy-terminal sequences that stimulate cGMP-dependent neuromodulation (see below).

To determine whether activation of NF- κ B caused an increase in gene expression, we tested the levels of interleukin-1 β (IL-1 β) and inducible nitric oxide synthase (iNOS) in the N9 cell line and in primary cultures of microglia. A 24-h treatment of primary microglia with 5 nM sAPP increased immunocytochemical staining for IL-1 β and iNOS (Fig. 2a); expression of IL-1 β and iNOS was also raised, as shown by western blot analysis of sAPP-treated microglia (Fig. 2b). The amounts of IL-1 β and iNOS in the N9 cell line responded to sAPP in a dose-dependent manner, with detectable induction at 100 pM (Fig. 2c). Increased expression of iNOS was apparent within 6 h of sAPP addition (data not shown).

It has been shown⁹ that the carboxy-terminal 15 residues that distinguish the products of α - and β -secretase activity (sAPP- α and sAPP- β , respectively) are critical for the cGMP-dependent neuroprotective effects of sAPP- α . We have confirmed these findings in a paradigm of neuronal death induced by 18 h of glucose deprivation of primary hippocampal neurons, in which treatment with 10 nM sAPP- α gave a survival rate of $222.3 \pm 19.6\%$ of control (neurons deprived only of glucose) and sAPP- β enhanced survival to only $126.5 \pm 12.0\%$ of control. However, such experiments have been done only in cultures lacking significant numbers of microglia. To test for any indirect effects of sAPP on neuronal survival mediated through microglia, we investigated whether there was a relation between specific sAPP sequences and microglial activation. As an index of activation, the culture medium of primary cultures of microglia was assayed for the presence of nitrite, a stable product of nitric oxide formation. Although there was a moderate difference, both sAPP- α and sAPP- β markedly increased nitrite concentrations in these cultures (Fig. 3). The neuroprotective effect of sAPP- α is retained by a construct containing residues 444–612 (β APP695 numbering)⁹; however, this construct was unable to activate micro-

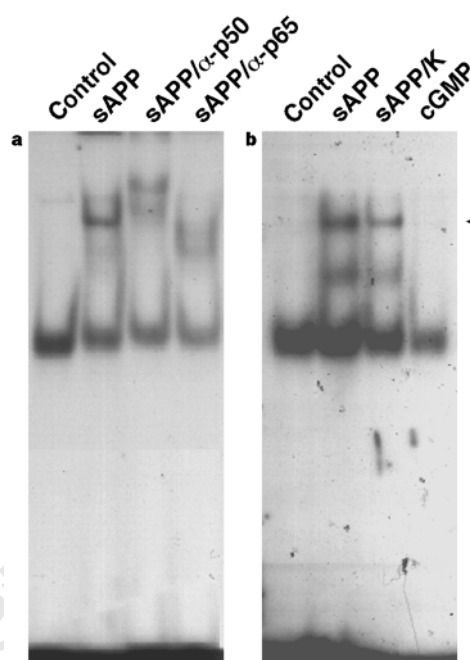


Figure 1 Activation of NF- κ B in N9 microglial cells by sAPP. N9 cultures were treated for 90 min with control buffer or 5 nM sAPP (sAPP), and nuclear extracts were assayed by EMSA with a κ B DNA probe. **a**, The sAPP-treated sample was also assayed after incubation with antibodies (Santa Cruz) generated against NF- κ B subunits p50 (sAPP/ α -p50) or p65 (sAPP/ α -p65). **b**, sAPP also was tested after a 10-min pretreatment of the cultures with 5 μ M KT5823, an inhibitor of the cGMP-dependent kinase (sAPP/K); 200 μ M 8-Br-cGMP was tested alone (cGMP). The arrow marks a specific, induced band that was either diminished or supershifted by NF- κ B antibodies.

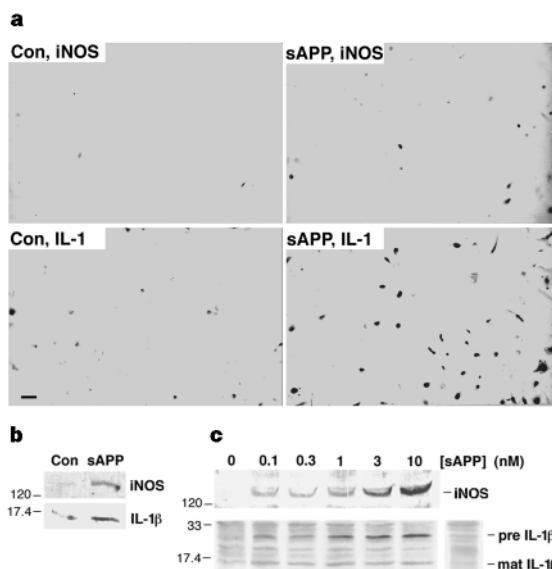


Figure 2 Increase in IL-1 β and iNOS in microglial cells by sAPP. **a**, **b**, Primary microglia were treated for 24 h with 5 nM sAPP (sAPP) or vehicle (Con) and fixed for immunocytochemistry (**a**) or collected for immunoblot analysis (**b**) with antibodies against iNOS or IL-1 β . Relative molecular mass (M_r) markers are shown on the left (in thousands). **c**, N9 microglial cells were treated with 5 nM sAPP for 16 h with the indicated concentrations of sAPP; cultures were harvested for immunoblot analysis with antibodies against iNOS (top; M_r ~130K) or IL-1 β (bottom). The lane of the far right shows the adjacent sample (10 nM-treated) probed with anti-IL-1 β that had been preincubated with purified IL-1 β . IL-1 β precursor (~30K) predominated over mature (mat) IL-1 β (~17K) in N9 lysates, whereas primary microglia appeared to have a much higher proportion of mature IL-1 β . M_r (K) markers are shown on the left.

glia efficiently (Fig. 3). These differences indicate that different domains of sAPP are involved in the protection of neurons and the activation of microglia, consistent with the lack of involvement of cGMP in activation. We next assayed this indirect neurotoxic effect of sAPP by applying soluble factors to microglia, then removing them before exposing the microglia to primary neurons. Primary

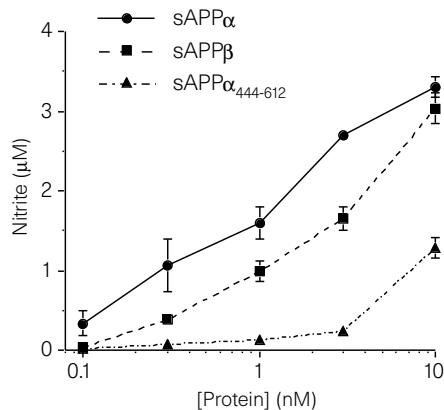


Figure 3 Structural requirements for sAPP stimulation of microglia. Constructs containing the coding sequences of human sAPP α , sAPP β and sAPP $\alpha_{444-612}$ were expressed in *E. coli* and purified to apparent homogeneity. Primary microglia were treated with 0.1–10 nM of each protein. After 24 h, the culture medium was tested for the presence of nitrite with Griess reagents. Values represent the mean $\pm$ s.e.m. for triplicate determinations in a single experiment representative of three performed. Results for sAPP $\alpha_{444-612}$ were significantly different from those for sAPP α and sAPP β ($P < 0.0003$ and $P < 0.04$, respectively); the difference between sAPP α and sAPP β was not significant.

microglia were plated onto a permeable membrane suspended in the bottom of a culture-well basket. Cultures were pretreated for 24 h with 5 nM sAPP α , sAPP β , or sAPP $\alpha_{444-612}$, then washed and transferred to 35-mm wells containing primary hippocampal neurons. Viability of all neurons decreased by $\sim 18\%$ over the next 48 h, and those exposed to untreated microglia were additionally compromised by a further 34% (Table 1). Toxicity was even greater in the presence of microglia that had been pretreated with sAPP α or sAPP β ; this correlation of microglial activation with neurotoxicity was extended to the reduced activity of sAPP $\alpha_{444-612}$ in both assays.

We have previously shown that the bioactivities of sAPP can be regulated differentially by two forms of human apolipoprotein E (ApoE) that are encoded by gene alleles in disequilibrium with Alzheimer's disease (AD). Specifically, interaction of ApoE3 with sAPP inhibits an activity associated with the amino-terminal 443 residues; ApoE4 was a less potent inhibitor¹⁰. Incubation with ApoE3 for 45 min also inhibited the ability of sAPP α to increase nitrite production in N9 cells (Fig. 4a) and to evoke microglia-mediated neurotoxicity (Fig. 4b); ApoE4 was less effective than

Table 1 Indirect neurotoxicity of sAPPs

	Neuronal survival (% of initial)	
	24 h	48 h
None (neurons alone)	93.1 $\pm$ 4.2	82.3 $\pm$ 6.7
Microglia, untreated	67.2 $\pm$ 9.1	48.0 $\pm$ 7.6
Microglia pretreated with sAPP α	37.4 $\pm$ 3.6*	16.9 $\pm$ 5.7*
Microglia pretreated with sAPP β	43.2 $\pm$ 5.0*	24.6 $\pm$ 4.9*
Microglia pretreated with sAPP $\alpha_{444-612}$	70.2 $\pm$ 5.4	39.4 $\pm$ 0.3

All sAPP pretreatments were applied at 3 nM. Neuronal survival was determined by a blinded observer using established morphological criteria from photographs of rat primary hippocampal neurons taken immediately before exposure to microglia ('initial') and at 24-h intervals thereafter. Data are expressed as the mean percentage ($\pm$ s.e.m.) of initial cell number present at each time point for triplicate cultures. * $P < 0.05$ versus second or fifth conditions (ANOVA and Scheffe post-hoc).

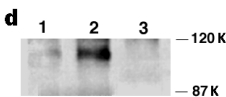
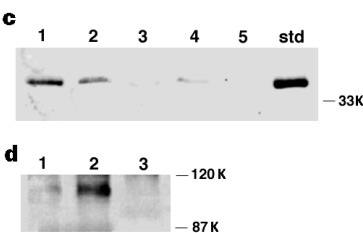
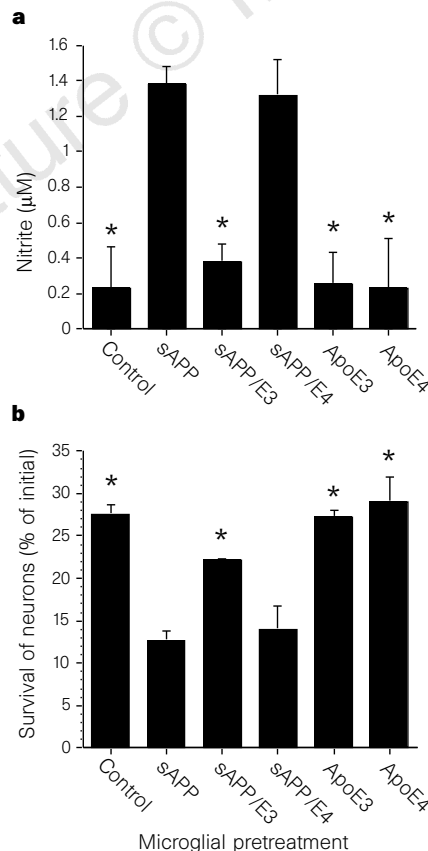


Figure 4 Interaction between ApoE and sAPP. **a**, Nitrite production in N9 cultures stimulated for 24 h with vehicle, 3 nM ApoE3, ApoE4 or sAPP α that had been incubated alone or with equimolar ApoE3 or ApoE4. **b**, Indirect neurotoxic effect on microglia treated as in **a** and assayed as for Table 1. **c**, ApoE3 (lanes 1, 3, 5) or ApoE4 (lanes 2, 4) was incubated alone (lanes 3, 4) or with equimolar polyhistidine-tagged sAPP α (lanes 1, 2) or sAPP $\alpha_{444-612}$ (lane 5). Precipitation with Ni²⁺-affinity resin was followed by immunoblotting with anti-ApoE (Boehringer-Mannheim); a sample of ApoE3 was included on the gel as a positive control (std). **d**, Mammalian-expressed sAPP α was incubated alone (lane 1) or with equimolar ApoE3 (lanes 2, 3). The mixture was precipitated with protein A (lane 3) or anti-ApoE protein A (lane 2), followed by immunoblotting with anti-APP. Positions of M_r markers kD are indicated on the right; * $P < 0.05$ versus sAPP.

ApoE3 in both respects. Co-precipitation experiments revealed that ApoE3 could bind sAPP- α (Fig. 4c) but not sAPP $\alpha_{444-612}$, suggesting that ApoE3 may bind and mask the sAPP domains responsible for activating microglia. The weaker binding of ApoE4 to sAPP- α (Fig. 4c) could explain why it does not affect sAPP- α activity.

Retrospective analysis¹¹ and preliminary clinical trials¹² have indicated that anti-inflammatory drugs may delay the onset or progression of AD. The crucial inflammatory events could involve microglia, which express markers of activation in AD. These activated microglia are clustered near amyloid plaques, particularly near subtypes of plaques that contain dystrophic neurites expressing high levels of β -APP¹³. Our data indicate that sAPP can activate microglia, enhancing their neurotoxicity. Considered together with the defective neuroprotection by sAPP- β , our results indicate that events favouring processing of β -APP by β -secretase might reduce the amounts of neuroprotective sAPP- α while retaining a detrimental activity mediated through microglia. The isoform-specific modulation by ApoE of detrimental microglial activation suggests that ApoE4 may allow more frequent neurodegenerative microglial reaction to insults, especially those that increase β -APP expression. For instance, β -APP increases after head injury (where AD or AD-like changes are synergistic with ApoE genotype^{14,15}), in epilepsy (where microglia also are activated¹⁶), and in normal ageing¹⁷⁻¹⁹ (when the interaction between ApoE4 and AD is most pronounced²⁰). Although our results suggest that the influence of ApoE results from its interaction with sAPP, direct immunomodulation by ApoE²¹ could explain effects on microglia. Amyloid- β (A β) can activate microglia²², an effect mediated through receptors for advanced-glycation end-products²³. The effects of A β on microglia show synergy with other cytokines²⁴, which might include sAPP. This suggests a role for sAPP even if it does not accumulate in AD brain, as indicated by measurements in cerebrospinal fluid^{25,26}. Our results do not exclude a contact-mediated effect on microglia by holo- β APP. It will be important pharmaceutically to identify the neurotoxin produced here: previous studies in which microglia were activated by amyloid plaques implicated a novel microglia-produced amine acting through glutamate receptors²⁷. □

Methods

Cells and reagents. Primary microglia were subcultured to ~95% purity from neonatal rat mixed glial cultures by differential adherence and panning techniques. Unless otherwise indicated, they were subcultured in minimal essential medium (MEM) supplemented to 10% with fetal bovine serum (FBS). The N9 cell line is a *myc*-immortalized murine microglial cell line generated by P. Ricciardi-Castagnoli; they were maintained in MEM/10% FBS and switched to serum-free MEM 18–24 h before stimulation. Primary cultures of hippocampal neurons were established from E18 rats as described⁸. Unless otherwise indicated, sAPP was purified (>98% homogeneity) from the conditioned medium of HEK293 transfectants as described⁸. The protein produced in this system is generated from β APP695 and has a C terminus consistent with α -secretase processing²⁸. Bacterially expressed protein for structural comparisons was generated from sequence coding for Val₂₀–Lys₆₁₂ of human β APP695 ('sAPP α ') placed in a pTrcHis (Invitrogen) expression vector. This vector tags expressed proteins N-terminally with a polyhistidine sequence to allow one-step purification on a nickel-affinity column. A second construct ('sAPP β ') was generated with a stop codon after Met 596. The third construct ('sAPP $\alpha_{444-612}$ ') was made from sAPP α by deletion of coding sequences aminoterminal to Asp 444. To exclude the possibility of contamination by bacterial endotoxin, some assays of these bacterially expressed proteins were done in the presence of 10 μ g ml⁻¹ polymyxin-B sulphate. Human recombinant ApoEs were obtained from Pan Vera and were not de-lipidated or exposed to reducing agents during purification. Antibodies included anti-iNOS monoclonal (Transduction Laboratories), hamster anti-murine IL-1 β monoclonal (Genzyme), and anti-ApoE monoclonal (Chemicon). Co-incubations of sAPP and ApoE were at room temperature for 45 min (polyhistidine-tagged sAPP) or 60 min (HEK293-expressed sAPP). For physiological assay, proteins were co-incubated at 30 nM each; for precipitation, the co-incubation

concentrations were 300 nM (polyhistidine-tagged sAPP) or 450 nM (HEK293-expressed sAPP). Precipitation reactions have been described¹⁰.

EMSA. Nuclear extracts were prepared according to ref. 29. 5 μ g extracted protein from each treatment condition was incubated with a ³²P-labelled, κ B DNA probe in EMSA buffer (50 mM Tris-HCl, pH 7.4, 20% glycerol, 50 mM NaCl, 5 mM MgCl₂, 2.5 mM EDTA, 0.5% Nonidet P-40, 5 mM β -mercaptoethanol, and 250 μ g ml⁻¹ poly(dI-dC)). Electrophoresis was done as described⁸.

Nitrite assay. For determination of nitrite, 100 μ l culture medium was removed and mixed with an equal volume of 0.5% sulphanilamide and 0.05% naphthylethylenediamine dihydrochloride in 0.25% phosphoric acid. After 10 min, the colour reaction was measured in a spectrophotometer at 540 nm and the readings calibrated from standards containing known amounts of nitrite. Data are presented as the mean $\pm$ s.e.m. for triplicate determinations within one of at least three similar experiments.

Statistics. Data were analysed by ANOVA with Scheffe *post hoc*, and *P*-values $\leq$ 0.05 were assumed to indicate significance.

Received 30 December 1996; accepted 12 June 1997.

- Hardy, J. New insights into the genetics of Alzheimer's disease. *Ann. Med.* **28**, 255–258 (1996).
- Rumble, B. *et al.* N. Engl. J. Med. **320**, 1446–1452 (1989).
- Griffin, W. S. *et al.* Microglial interleukin-1 α expression in human head injury: correlations with neuronal and neuritic β -amyloid precursor protein expression. *Neurosci. Lett.* **176**, 133–136 (1994).
- Moran, P. M., Higgins, L. S., Cordell, B. & Moser, P. C. Age-related learning deficits in transgenic mice expressing the 751-amino acid isoform of human β -amyloid precursor protein. *Proc. Natl Acad. Sci. USA* **92**, 5341–5345 (1995).
- Mrak, R. E., Sheng, J. G. & Griffin, W. S. Glial cytokines in Alzheimer's disease: review and pathogenic implications. *Hum. Pathol.* **26**, 816–823 (1995).
- Roses, A. D. Apolipoprotein E alleles as risk factors in Alzheimer's disease. *Annu. Rev. Med.* **47**, 387–400 (1996).
- Corradin, S. B., Muel, J., Donini, S. D., Quattrocchi, E. & Ricciardi-Castagnoli, P. Inducible nitric oxide synthase activity of cloned murine microglial cells. *Glia* **7**, 255–262 (1993).
- Barger, S. W. & Mattson, M. P. Induction of neuroprotective κ B-dependent transcription by secreted form of the Alzheimer's β -amyloid precursor. *Mol. Brain Res.* **40**, 116–126 (1996).
- Furukawa, K. *et al.* Increased activity-regulating and neuroprotective efficacy of α -secretase-derived secreted amyloid precursor protein conferred by a c-terminal heparin-binding domain. *J. Neurochem.* **67**, 1882–1896 (1996).
- Barger, S. W. & Mattson, M. P. Isoform-specific modulation by apolipoprotein E of the activities of secreted β -amyloid precursor protein. *J. Neurochem.* **69**, 60–67 (1997).
- Breitner, J. C. *et al.* Delayed onset of Alzheimer's disease with nonsteroidal anti-inflammatory and histamine H2 blocking drugs. *Neurobiol. Aging* **16**, 523–530 (1995).
- Rogers, J. *et al.* Clinical trial of indomethacin in Alzheimer's disease. *Neurology* **43**, 1609–1611 (1993).
- Griffin, W. S., Sheng, J. G., Roberts, G. W. & Mrak, R. E. Interleukin-1 expression in different plaque types in Alzheimer's disease: significance in plaque evolution. *J. Neuropathol. Exp. Neurol.* **54**, 276–281 (1995).
- Mayeux, R. *et al.* Synergistic effects of traumatic head injury and apolipoprotein-e4 in patients with Alzheimer's disease. *Neurology* **45**, 555–557 (1995).
- Nicoll, J. R., Roberts, G. W. & Graham, D. I. Apolipoprotein E ϵ 4 allele is associated with deposition of amyloid β -protein following head injury. *Nat. Med.* **1**, 135–137 (1995).
- Sheng, J. G., Boop, F. A., Mrak, R. E. & Griffin, W. S. Increased neuronal β -amyloid precursor protein expression in human temporal lobe epilepsy: association with interleukin-1 α immunoreactivity. *J. Neurochem.* **63**, 1872–1879 (1994).
- Nordstedt, C. *et al.* Alzheimer β /A4 amyloid precursor protein in human brain: aging-associated increases in holoprotein and in a proteolytic fragment. *Proc. Natl Acad. Sci. USA* **88**, 8910–8914 (1991).
- Martin, L. J., Pardo, C. A., Cork, L. C. & Price, D. L. Synaptic pathology and glial responses to neuronal injury precede the formation of senile plaques and amyloid deposits in the aging cerebral cortex. *Am. J. Pathol.* **145**, 1358–1381 (1994).
- Loffler, K. U., Edward, D. P. & Tso, M. O. Immunoreactivity against tau, amyloid precursor protein, and β -amyloid in the human retina. *Invest. Ophthalmol. Vis. Sci.* **36**, 24–31 (1995).
- Corder, E. H. *et al.* Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science* **261**, 921–923 (1993).
- Clay, M. A., Anantharamaiah, G. M., Mistry, M. J., Balasubramanian, A. & Harmony, J. A. Localization of a domain in apolipoprotein E with both cytostatic and cytotoxic activity. *Biochemistry* **34**, 11142–11151 (1995).
- Meda, L. *et al.* Activation of microglial cells by β -amyloid protein and interferon- γ . *Nature* **374**, 647–650 (1995).
- Yan, S. D. *et al.* RAGE and amyloid- β peptide neurotoxicity in Alzheimer's disease. *Nature* **382**, 685–691 (1996).
- Rossi, F. & Bianchini, E. Synergistic induction of nitric oxide by β -amyloid and cytokines in astrocytes. *Biochem. Biophys. Res. Commun.* **225**, 474–478 (1996).
- Palmer, M. R. *et al.* Soluble derivatives of the β amyloid protein precursor in cerebrospinal fluid: alterations in normal aging and in Alzheimer's disease. *Neurology* **40**, 1028–1034 (1990).
- Van Nostrand, W. E. *et al.* Decreased levels of soluble amyloid β -protein precursor in cerebrospinal fluid of live Alzheimer disease patients. *Proc. Natl Acad. Sci. USA* **89**, 2551–2555 (1992).
- Giulian, D. *et al.* Senile plaques stimulate microglia to release a neurotoxin found in Alzheimer brain. *Neurochem. Int.* **27**, 119–137 (1995).
- Oltersdorf, T. *et al.* The secreted form of the Alzheimer's amyloid precursor protein with the Kunitz domain is protease nexin-II. *Nature* **341**, 144–147 (1989).
- Ostrowski, J. *et al.* A serine/threonine kinase activity is closely associated with a 65-kDa phosphoprotein specifically recognized by the κ B enhancer element. *J. Biol. Chem.* **266**, 12722–12733 (1991).

Acknowledgements. We are indebted to Paola Ricciardi-Castagnoli for the N9 cell line, to D. Davies for advice on microglial culture techniques, to M. Lucas for technical assistance, and to S. Griffin and M. J. LaDu for helpful comments. This work was supported by funds granted to S.W.B. by the Ingelwood Foundation, NIH (NINDS), Alzheimer's Association, and a UAMS pilot project award.

Correspondence and requests for materials should be addressed to S.W.B. (e-mail: swbarger@life.uams.edu).

Fluorescent indicators for Ca^{2+} based on green fluorescent proteins and calmodulin

Atsushi Miyawaki*, Juan Llopis*, Roger Heim*†, J. Michael McCaffery‡, Joseph A. Adams§, Mitsuhiro Ikura||, & Roger Y. Tsien*†

* Department of Pharmacology, † Howard Hughes Medical Institute, and

‡ Division of Cellular and Molecular Medicine, University of California, San Diego, La Jolla, California 92093-0647, USA

§ Department of Chemistry, San Diego State University, San Diego, California 92182-1030, USA

|| Division of Molecular and Structural Biology, Ontario Cancer Institute, and Department of Medical Biophysics, University of Toronto, Toronto M5G 2M9, Canada, and Center for Tsukuba Advanced Research Alliance, University of Tsukuba, Tsukuba 305, Japan

Important Ca^{2+} signals in the cytosol and organelles are often extremely localized and hard to measure. To overcome this problem we have constructed new fluorescent indicators for Ca^{2+} that are genetically encoded without cofactors and are targetable to specific intracellular locations. We have dubbed these fluorescent indicators 'cameleons'. They consist of tandem fusions of a blue- or cyan-emitting mutant of the green fluorescent protein (GFP)^{1,2}, calmodulin³⁻⁵, the calmodulin-binding peptide M13 (ref. 6), and an enhanced green- or yellow-emitting GFP⁷⁻⁹. Binding of Ca^{2+} makes calmodulin wrap around the M13 domain, increasing the fluorescence resonance energy transfer

(FRET) between the flanking GFPs². Calmodulin mutations can tune the Ca^{2+} affinities to measure free Ca^{2+} concentrations in the range 10^{-8} to 10^{-2} M. We have visualized free Ca^{2+} dynamics in the cytosol, nucleus and endoplasmic reticulum of single HeLa cells transfected with complementary DNAs encoding chimaeras bearing appropriate localization signals. Ca^{2+} concentration in the endoplasmic reticulum of individual cells ranged from 60 to 400 μM at rest, and 1 to 50 μM after Ca^{2+} mobilization. FRET is also an indicator of the reversible intermolecular association of cyan-GFP-labelled calmodulin with yellow-GFP-labelled M13. Thus FRET between GFP mutants can monitor localized Ca^{2+} signals and protein heterodimerization in individual live cells.

Cytosolic and organellar free Ca^{2+} concentrations are among the most important and dynamic intracellular signals, and are usually measured using synthetic fluorescent chelators¹⁰⁻¹³ or recombinant aequorin^{14,15}. The chelators are easily imaged but are difficult to target precisely to specific intracellular locations, whereas aequorin is easily targeted but requires the incorporation of coelenterazine, is irreversibly consumed by Ca^{2+} , and is very difficult to image because its luminescence produces less than one photon per molecule. To combine the brightness of fluorescent indicators with the targetability of a biosynthetic indicator, we have used GFP, a spontaneously fluorescent protein from the jellyfish *Aequorea victoria*. Its cDNA can be concatenated with those encoding many other proteins, and the resulting fusion proteins are usually fluorescent and often preserve the biochemical functions and cellular localizations of the partner proteins. Mutagenesis has produced GFP mutants with shifted wavelengths of excitation or emission^{1,2,7,9} that can serve as donors and acceptors for FRET. FRET is a non-destructive spectroscopic method that can monitor the proximity and relative angular orientation of fluorophores in living cells. The donor and acceptor fluorophores can be entirely separate or

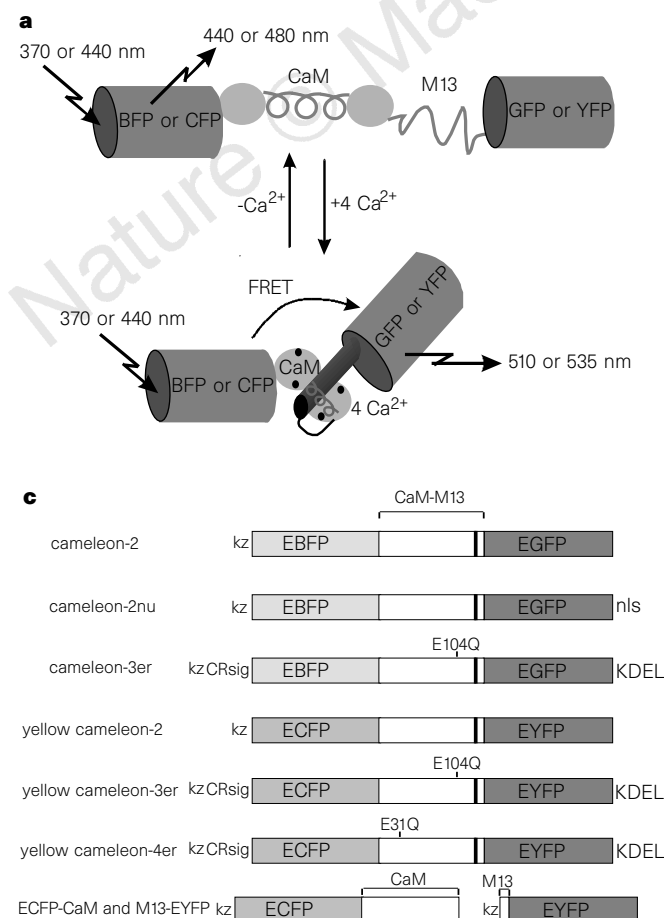


Figure 1 Schematic structures and sequences of cameleons. **a**, Scheme showing how FRET between GFPs can measure Ca^{2+} . The GFPs are drawn as simple rigid cylinders, reflecting their crystal structures⁹. Schematic structures of calmodulin (CaM) without Ca^{2+} , disordered unbound M13, and the Ca^{2+} -calmodulin-M13 complex are derived from crystallography⁴, NMR⁶ and modelling¹⁶, but the relative orientations of the GFPs are unknown. **b**, Domain structure of cameleons expressed in bacteria for *in vitro* characterization, showing sequences of the boundaries between the donor GFP and *Xenopus* calmodulin (XCaM) and between M13 and the acceptor GFP. **c**, Construction of cameleons for expression and imaging in mammalian cells. CRsig, calreticulin signal sequence, MLLSVPLLLGLLGLAAAD²⁴; nls, nuclear localization signal, PKKKRKVEDA; KDEL, ER retention signal²⁴; kz, Kozak consensus sequence for optimal translational initiation in mammalian cells²⁸. The significant mutations in the four GFPs are: EBFP, F64L/Y66H/Y145²; EGFP, F64L/S65T⁹; ECFP, F64L/S65T/Y66W/N146I/M153T/V163A/N212K²; and EYFP, S65G/S72A/T203Y⁹.

attached to the same macromolecule. In the latter case, ligand-induced conformational changes might be monitored by FRET if the amino and carboxy termini of the binding protein are fused to a donor and acceptor GFP. This approach has several advantages over the usual covalent labelling with fluorescent probes. First, the indicator is generated *in situ* by gene transfer into the cells or organisms, obviating large-scale expression, purification, labelling and microinjection of recombinant proteins that must be soluble. Second, the sites of fusions are defined exactly, giving a molecularly homogenous product without the use of elaborate protein chemistry. Third, the chromophore of GFP is fixed in the protein⁹. If the GFP donor and acceptor are rigidly fused to a host protein, minor changes in the relative orientation of the ends of the latter would alter FRET. In contrast, most conventional fluorescent labels are attached by flexible linkers that at least partly decouple the fluorophore orientation from that of the protein to which it is attached.

Many effects of Ca^{2+} in cells are mediated by the binding of Ca^{2+} to calmodulin, which causes calmodulin to bind and activate target proteins³. Based on the nuclear magnetic resonance solution structure of calmodulin bound to M13, the 26-residue calmodulin-binding peptide of myosin light-chain kinase⁶, the C terminus of calmodulin has been fused to the M13 by means of a Gly-Gly spacer¹⁶. Ca^{2+} binding switched the resulting hybrid protein (calmodulin-M13) from a dumb-bell-like extended form to a compact globular form similar to the calmodulin-M13 intermolecular complex. We therefore sandwiched the calmodulin-M13 fusion between a blue (Y66H/Y145F)² and a green (S65T)⁷ GFP mutant to construct a Ca^{2+} indicator (Fig. 1a). The amino-acid sequences of

the boundary regions between the calmodulin-M13 hybrid and GFPs proved critical to the optimization of protein folding and the Ca^{2+} sensitivity of FRET. Numerous deletions, insertions and amino-acid substitutions were tested; the best splice sequences found are shown in Fig. 1b. The prototype fusion protein was efficiently expressed and folded in bacteria and increased its ratio of ultraviolet-excited 510:445 nm emissions by 70% upon binding Ca^{2+} (Fig. 2a). The decrease in blue and increase in green emission indicated that Ca^{2+} increased the efficiency of FRET from Y66H/Y145F to S65T, consistent with the expected decrease in distance between the two ends of the protein. The Ca^{2+} response was fully reversible upon chelation of Ca^{2+} . We have called this calmodulin-based indicator 'cameleon-1', because it readily changes colour and retracts and extends a long tongue (M13) into and out of the mouth of the calmodulin, often abbreviated CaM. Mg^{2+} , pH and ionic strength did not alter the emission spectra of either the Ca^{2+} -saturated or Ca^{2+} -unsaturated forms (data not shown), in accord with the typical high specificity of calmodulin for Ca^{2+} (ref. 5). Cameleon-1 was not affected by hydrophobic proteins, such as bovine serum albumin. Isolated calmodulin saturated with Ca^{2+} becomes sticky with hydrophobic amino acids exposed to the surface. The calmodulin in cameleon-1, however, seems to interact preferentially with its intramolecularly adjacent M13 peptide. Such a self-contained system should minimize the possibility that the protein might interact with endogenous calmodulin-binding sequences in eukaryotic cells.

The original calmodulin-M13 hybrid protein without GFPs displayed a biphasic Ca^{2+} binding with two dissociation constants,

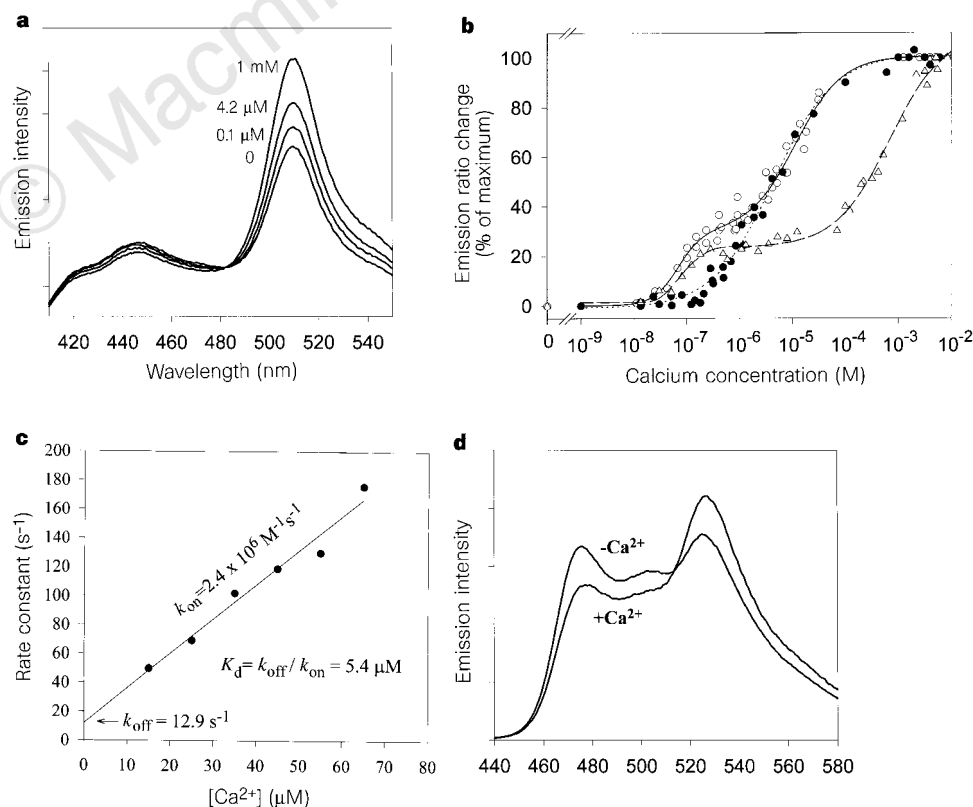


Figure 2 Properties of cameleons *in vitro*. **a**, Emission spectra of cameleon-1 (excited at 380 nm) at the indicated free Ca^{2+} values at pH 7.47. **b**, Ca^{2+} titration curves of cameleon-1 (open circles), cameleon-1/E104Q (filled circles), and cameleon-1/E31Q (open triangles), combining 5 independent experiments using Ca^{2+} /EGTA and Ca^{2+} /HEEDTA systems¹⁰ below 10^{-6} M free Ca^{2+} and unbuffered Ca^{2+} above. The changes in emission ratio (510 to 445 nm) were normalized to the

effects of full Ca^{2+} saturation, which were 60–80% increases in ratio over the values at zero Ca^{2+} . The fitted curves correspond to the apparent dissociation constants and Hill coefficients given in the text. **c**, Relaxation rate constants, k_{obs} ($= k_{\text{on}}[\text{Ca}^{2+}] + k_{\text{off}}$), for reaction of cameleon-1/E104Q with Ca^{2+} ; k_{on} and k_{off} are the association and dissociation rate constants. **d**, Emission spectrum of yellow cameleon-2 *in vitro*, excited at 432 nm, at zero and saturating Ca^{2+} .

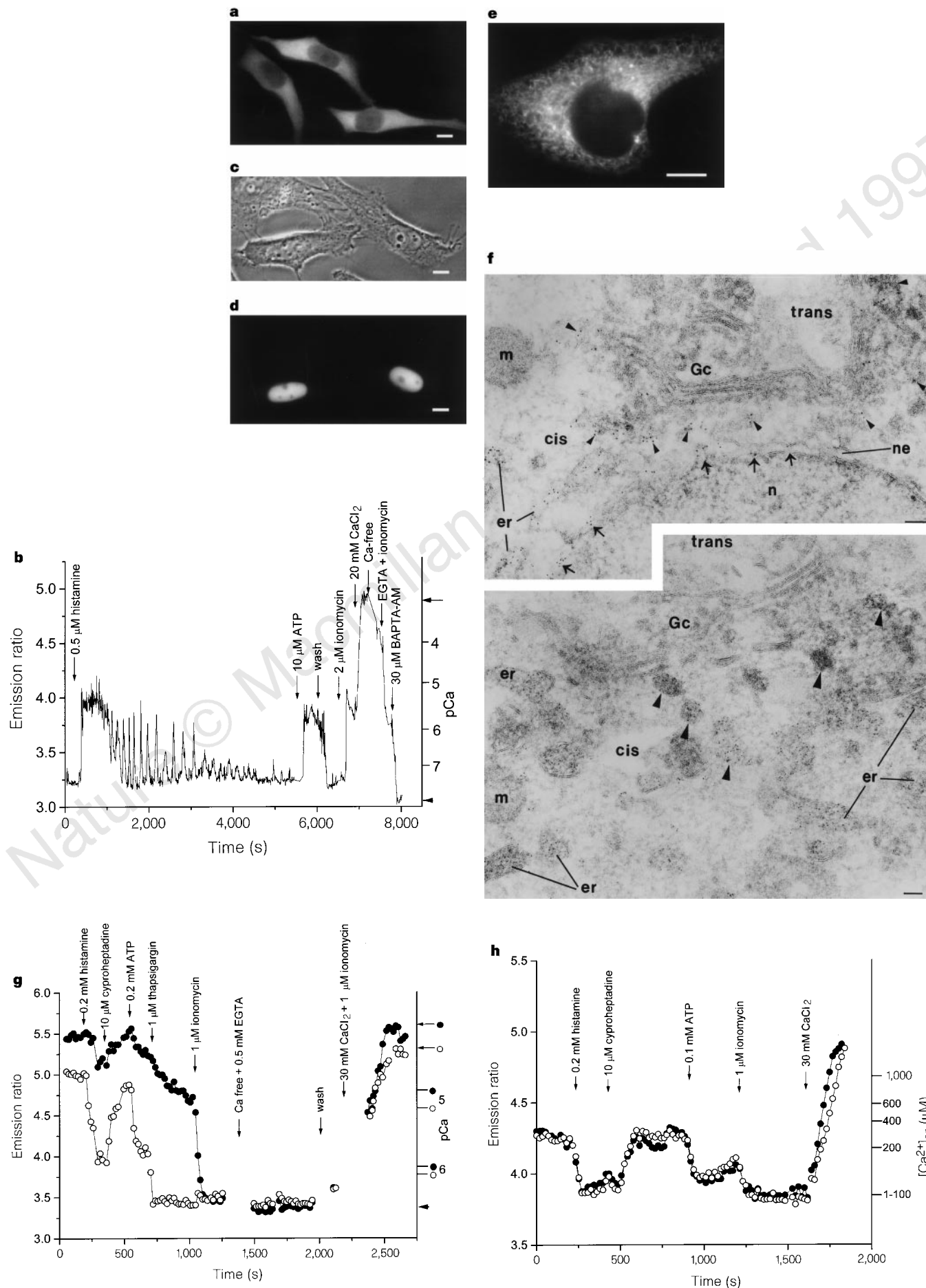


Figure 3 Imaging of cameleons in HeLa cells. **a**, Fluorescence image (440 ± 10 nm excitation, 535 ± 12.5 nm emission) of yellow cameleon-2 shows cytosolic localization in HeLa cells. Scale bar, $10 \mu\text{m}$ in all fluorescence images. **b**, Ratios of 535 ± 12.5 nm to 480 ± 15 nm emissions from yellow cameleon-2 in the cytoplasm of the leftmost cell of **a**, monitored every 7 s by digital imaging microscopy. The ratios were calibrated in terms of absolute pCa on the right-hand ordinate axis, using $R_{\min}$ (arrowhead) and $R_{\max}$ (arrow) determined *in situ*. 'Ca²⁺ free' indicates medium containing 0.5 mM MgCl₂. **c, d**, Matched phase-contrast and fluorescence (480 ± 15 nm excitation, 535 ± 22.5 nm emission) images of two HeLa cells showing cameleon-2nu localized to the nuclei. **e**, Fluorescence image of yellow cameleon-3er, taken as in **a**. The same reticular pattern of fluorescence was observed for yellow cameleon-4er. **f**, Immunogold localization of cameleon-3er in ultrathin cryosections using a polyclonal antibody against GFP. The cameleon-3er was abundant in the ER (er), nuclear envelope (ne and arrows) and intermediate compartment (arrowheads), but absent from the *cis*-most Golgi cisternae (Gc), mitochondria (m) and nucleus (n). 'Cis' and 'trans' show the orientation of the Golgi stack but do not mark specific structures. Scale bars, $0.1 \mu\text{m}$. **g**, The 535 to 480 nm emission ratios from yellow cameleon-3er in two cells. The right-hand ordinate gives $R_{\min}$ (arrowhead), $R_{\max}$ (arrows) and pCa calibrations for each cell. **h**, Analogous emission ratios from yellow cameleon-4er in two cells. The right-hand ordinate calibrates $[\text{Ca}^{2+}]_{\text{er}}$.

2 μM and 80 nM, resulting from independent Ca²⁺ binding to the N- and C-terminal domains of calmodulin, respectively¹⁶. The emission ratio of cameleon-1 has a similar biphasic Ca²⁺ dependency (Fig. 2b, open circles) with apparent dissociation constants (K'_d) of 11 μM and 70 nM, and Hill coefficients (n) of 1.0 and 1.8, respectively. Therefore cameleon-1 can report a very wide range of Ca²⁺ concentration, from $<10^{-7}$ to $>10^{-4}$ M. Many site-directed mutations have been studied for their effects on the Ca²⁺ binding and Ca²⁺-induced conformational changes of calmodulin^{17,18}. Mutation of the conserved bidentate glutamic acid (E104) at position 12 of the third Ca²⁺-binding loop to glutamine (Q) eliminated the high-affinity component of the Ca²⁺ response (Fig. 2b, filled circles). The resulting monophasic response (K'_d , 4.4 μM ; n , 0.76) corresponded fairly closely to the low-affinity component of the cameleon-1 response. An analogous mutation in the first Ca²⁺-binding loop (E31Q) resulted in a large rightward shift of the low-affinity component of cameleon-1, without affecting the high-affinity component (Fig. 2b, triangles). Cameleon-1/E31Q showed a biphasic response (K'_d , 83 nM and 700 μM ; n , 1.5 and 0.87. Much more tuning of Ca²⁺ affinities for particular applications should be possible.

Because cameleon-1/E104Q had the simplest Ca²⁺-response curve, it was chosen for initial explorations of Ca²⁺-binding kinetics by stopped-flow measurements of the acceptor S65T emission after rapid mixing with Ca²⁺. The approach to equilibrium was reasonably exponential, and the inverse time constants were a linear function of free $[\text{Ca}^{2+}]$, as expected for simple 1:1 binding (Fig. 2c). The association and dissociation rate constants correspond to the slope and intercept of the linear plot and were $(2.4 \pm 0.2) \times 10^6 \text{ M}^{-1} \text{ s}^{-1}$ and $13 \pm 9 \text{ s}^{-1}$, in 100 mM KCl at 25 °C. Their ratio is 5.4 μM , which is consistent with the apparent dissociation constant (4.4 μM) independently derived from equilibrium measurements (Fig. 2b). The dissociation rate is similar to that ($12 \pm 2 \text{ s}^{-1}$) for calmodulin and M13 without the GFPs¹⁹, so the GFPs do not perturb the dissociation even though they are much larger than the interacting calmodulin and M13 domains.

Can cameleons work as Ca²⁺ indicators in mammalian cells? Initial transfections with the gene encoding cameleon-1 in the pcDNA3 vector (with a cytomegalovirus promoter) gave inadequate fluorescence because the blue was not bright enough. The simplest remedy was to introduce mammalian codon bias and mutate Phe 64 to Leu, which improves expression and folding of

GFPs at 37 °C without greatly affecting the spectra of the final well-folded molecules^{8,20}. The resulting enhanced blue and green fluorescent proteins ('EBFP' and EGFP, respectively) fused to calmodulin-M13 constituted cameleon-2 (Fig. 1c), which indeed greatly improved mammalian cell expression. However, an even better approach for single-cell imaging proved to be replacement of the EBFP and EGFP by similarly enhanced cyan and yellow fluorescent mutants, 'ECFP' and 'EYFP', respectively (Fig. 1c). ECFP contains Trp at residue 66 in place of the His of EBFP and Tyr of EGFP, whereas EYFP contains the normal Tyr 66 in the fluorophore but is shifted to a yellowish emission by mutation of Thr 203 to Tyr (ref. 9). Both genes have mammalian codon usage and encode appropriate mutations to improve folding²⁰. The substitution of ECFP for EBFP shifted the excitation peak from 381 to 433 nm and improved the brightness, signal-to-noise ratio and duration of recording, yet decreased potential concerns regarding autofluorescence, photo-damage and incompatibility with caged compounds. The 'yellow cameleons' incorporating ECFP and EYFP have three minor disadvantages. The maximal ratio changes (Fig. 2d) are slightly smaller (~1.5-fold) than for the original EBFP/EGFP cameleons (1.8-fold), perhaps because the emission spectrum of ECFP encroaches more on the EYFP emission. Yellow cameleons are perturbed by acidification below pH 7.0, which mimics a fall in Ca²⁺. Also, EYFP can undergo reversible photochromism if illuminated too strongly (A. Miyawaki and J. Llopis, unpublished observations). Other properties, such as Ca²⁺ affinities, as well as the qualitative behaviour inside cells, seem to be the same for the two families of cameleons.

When yellow cameleon-2 was transfected into HeLa cells, the fluorescence was uniformly distributed in the cytosolic compartment but excluded from the nucleus (Fig. 3a), as expected for a 74 kDa protein without targeting signals. The time course of the spatially averaged yellow: cyan emission ratios from a single HeLa cell expressing yellow cameleon-2 is shown in Fig. 3b. A submaximal dose of histamine (0.5 μM) produced a significant increase in the emission ratio indicating an initial peak of about 3 μM cytosolic free Ca²⁺, $[\text{Ca}^{2+}]_c$. This response was gradually converted into oscillations and eventually desensitized altogether. However, ATP added as a purinergic agonist could still stimulate another increase in $[\text{Ca}^{2+}]_c$ of similar amplitude to the initial histamine response. These responses agree well with the known behaviour of HeLa cells to histamine and ATP^{21,22}. Application of the Ca²⁺ ionophore ionomycin followed by a high concentration (20 mM) of extracellular Ca²⁺ gave a large increase in the ratio, which should correspond to the maximal ratio $R_{\max}$. To establish the minimal ratio $R_{\min}$, as required for a simple *in situ* calibration, the cell was loaded with the permeant chelator ester BAPTA/AM, and external Ca²⁺ was clamped to zero with EGTA in the presence of ionomycin. The ratio $R_{\max}/R_{\min}$ inside cells was about 1.6, equal or slightly greater than that observed *in vitro*. This correspondence argues that the indicator in cells behaved in reasonable accord with its *in vitro* calibration, and that the calibration protocol effectively achieved saturating and zero $[\text{Ca}^{2+}]_c$.

Cameleons should be readily directable to interesting sites by fusion to appropriate organellar targeting signals or localized host proteins. For example, addition of a nuclear localization signal to cameleon-2 yielded a Ca²⁺ indicator, 'cameleon-2nu' (Fig. 1c), fluorescence of which was tightly localized to nuclei (Fig. 3c, d) but excluded from nucleoli. The time course of nuclear $[\text{Ca}^{2+}]$ proved generally similar to those of $[\text{Ca}^{2+}]_c$, confirming previous comparisons of nuclear and cytosolic aequorin²³. We also succeeded in monitoring agonist-induced changes in the free Ca²⁺ inside the endoplasmic reticulum ($[\text{Ca}^{2+}]_{\text{er}}$) in individual, intact, non-perfused cells. Two low-affinity indicators, yellow cameleon-3er and 4er (for endoplasmic reticulum) (Fig. 1c), whose Ca²⁺ responses should correspond to the filled circles and open triangles respectively of Fig. 2b, were engineered to reside in the lumen of

endoplasmic reticulum (ER) by addition of a signal sequence at the N terminus and a KDEL signal for ER retention at the C terminus²⁴. Reticular patterns of fluorescence were seen in HeLa cells expressing the protein (Fig. 3e). Targeting to the ER was verified by electron-microscopic immunogold localization (Fig. 3f), which showed that almost all of the GFP immunoreactivity was within the ER and nuclear envelope. A few percent of gold particles were in the intermediate compartment between ER and Golgi, but the Golgi apparatus itself was essentially devoid of immunoreactivity. $[Ca^{2+}]_{er}$ reported by yellowameleon-3er in two neighbouring cells is shown in Fig. 3g. In one cell (filled circles), the indicator was initially saturated ($[Ca^{2+}]_{er} > 100 \mu M$), as shown by an emission ratio matching R_{max} and a time delay in response to the Ca^{2+} -mobilizing effect of histamine. The agonist eventually depleted $[Ca^{2+}]_{er}$ to about $30 \mu M$, within the dynamic range of the indicator. In the other cell (open circles), $[Ca^{2+}]_{er}$ started at about $60 \mu M$ and responded to histamine with much less time delay, soon levelling off at $1.5 \mu M$. The depletion of $[Ca^{2+}]_{er}$ was promptly reversed by a histamine antagonist, cyproheptadine, and recapitulated by ATP. Responses to thapsigargin blockage of the ER Ca^{2+} -ATPase were also heterogeneous, whereas ionomycin immediately clamped $[Ca^{2+}]_{er}$ almost to zero. To measure resting $[Ca^{2+}]_{er}$ values above $100 \mu M$, we turned to the lowest-affinity indicator, yellowameleon-4er (Figs 2b and 3h). This variant was clearly not saturated, because the ratio started well below R_{max} , fell immediately upon histamine stimulation, but was delayed in recovering after antagonist application. Combined results from the two indicators separately applied to 75 cells showed $[Ca^{2+}]_{er}$ ranging from 60 to $400 \mu M$ before and 1 to $50 \mu M$ after application of a maximal dose of agonists. Our estimates for basal $[Ca^{2+}]_{er}$ fall roughly in the middle of previous estimates, which range from 1 to $2,000 \mu M$ (refs 11, 14, 15, 25). Previous measurements using small-molecule indicators are subject to severe uncertainties in dye localization and Mg^{2+} interference. Aequorin experiments can only be done on cell populations and are complicated by the opposing abilities of Ca^{2+} both to stimulate light output and to irreversibly destroy the indicator. ER-targeted cameleons avoid both sets of problems. Ca^{2+} stores in the ER have recently been suggested to contain spatially and functionally distinct

compartments^{13,26}. Cameleons fused to different ER sequences should be able to examine such heterogeneity with the unique combination of precise molecular targeting, single-cell physiological imaging, and ultrastructural localization.

Is FRET between GFPs limited to detecting intramolecular proximity, or can it also detect intermolecular associations? Yellowameleon-2 was expressed as two separate pieces: ECFP-calmodulin and M13-EYFP (Fig. 1c). When these two proteins were mixed *in vitro*, the spectral difference between zero and saturating Ca^{2+} was similar to but somewhat larger than that for the corresponding intramolecular association in the fully fused construct (data not shown), presumably because the unfused partners could get much further apart from each other in the absence of Ca^{2+} . When expressed in HeLa cells, ECFP-calmodulin and M13-EYFP were distributed throughout both the cytoplasm and the nucleus, as would be expected from their molecular weights (M_r 44K and 30K, respectively). This pattern is quite different from the exclusion of intact yellowameleon-2 from the nucleus. Histamine mobilization of Ca^{2+} elevated the emission ratio of EYFP to ECFP in both cytoplasm and nucleus, an effect reversed by a histamine antagonist (Fig. 4, circles). ATP caused a similarly reversible increase in the ratio. Such physiological $[Ca^{2+}]_c$ transients caused reciprocal changes in the EYFP and ECFP intensities (Fig. 4, lines without symbols), as expected for association/dissociation effects on FRET, whereas artefacts such as a change of focus caused parallel changes that were cancelled by calculating the ratios. Thus GFP-based FRET can monitor the dynamics of reversible heterodimerization in live cells. In a recent alternative approach²⁷, blue- and green-emitting GFPs were linked through the calmodulin-binding domain of smooth-muscle myosin light-chain kinase. FRET between the GFPs was disrupted *in vitro* upon binding of unlabelled $(Ca^{2+})_4$ -calmodulin to the linker. The fusion protein was introduced into cells only by microinjection, and showed small fluorescence response to $[Ca^{2+}]_c$ that were amplifiable by co-injection of exogenous calmodulin.

Many further improvements in cameleons should be possible by mutagenesis, whereas small-molecule indicators can only be optimised by more laborious chemical resynthesis. The targetability of

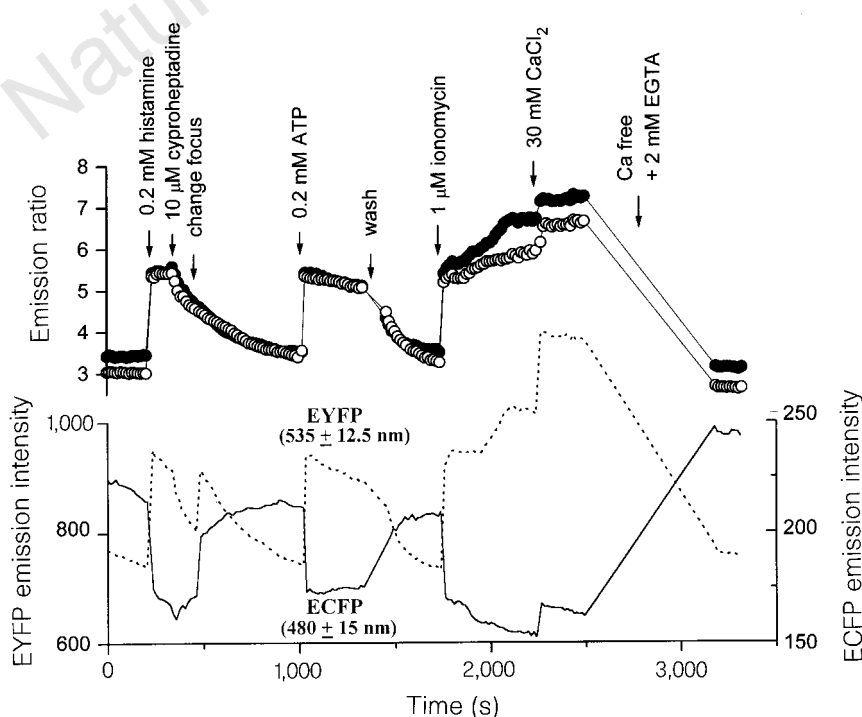


Figure 4 GFP-based detection of protein association/dissociation. Top, 535 to 480 nm emission ratios (filled and open circles) in two adjacent cells co-transfected with M13-EYFP and ECFP-calmodulin. Bottom, 535 (broken line, left-hand scale) and 480 nm (solid line, right-hand scale) emissions, background-corrected, from the cell corresponding to the filled circles.

cameleons may eventually allow Ca^{2+} measurements at previously inaccessible sites, such as the immediate vicinity of synaptic vesicles or Ca^{2+} channels. Cameleons should be well suited to genetically tractable organisms such as bacteria, yeast, nematodes, flies, plants and transgenic mice. Fluorescent indicators for analytes other than Ca^{2+} might be engineered by sandwiching other conformationally responsive receptors between GFP donors and acceptors. Fluorescence readout of intermolecular associations by FRET (as in Fig. 4) is non-destructive, quantifiable with high spatiotemporal resolution, and probably applicable to any compartment of the cell and to many proteins other than calmodulin and M13. However, the ratio of intensities in the acceptor and donor channels starts from a non-zero baseline even in the absence of FRET. Therefore, rare associations will be hard to detect, and positive and negative controls are crucial, such as the ability to switch the putative association on and off, ideally *in situ*. The dynamic range of such measurements both of Ca^{2+} and of protein association will benefit from further improvements in the GFP mutants to reduce leakage of the donor emission into the acceptor emission band. □

Methods

Gene construction. The cDNA of the GFP blue mutant Y66H/Y145F (ref. 1) was amplified by the polymerase chain reaction (PCR) with a sense primer containing a *Bam*HI site and a reverse primer containing an *Sph*I site and eliminating the GFP stop codon. Similarly the cDNA of S65T was amplified with a *Sac*I site and an *Eco*RI site introduced to the 5' and 3' ends of the gene, respectively. Two restriction sites (*Sph*I and *Sac*I) were introduced by PCR into the 5' and 3' ends of the calmodulin-M13 gene, respectively, using pHY1 (ref. 16) as a template. The restricted products were ligated and cloned in-frame into the *Bam*HI/*Eco*RI sites of pRSETB (Invitrogen). The modifications of the boundary regions between Y66H/Y145F and calmodulin and between M13 and S65T were performed by PCR or by a combined use of restriction enzymes, Klenow fragment of DNA polymerase I, T4 DNA polymerase, mung-bean exonuclease, and T4 DNA ligase. Enhanced GFP mutants EBFP, ECFP and EYFP were created by introducing appropriate substitutions^{2,9} into the commercially available pEGFP plasmid from Clontech, which has mammalian codon usage²⁰ and mutations F64L and S65T. Oligonucleotide-directed mutageneses were performed using the Muta-Gene Phagemid *in vitro* kit (Bio-Rad). The 5' end of the EBFP or ECFP gene was modified by PCR to have a *Hind*III restriction site followed by a Kozak consensus sequence²⁸ (ACCGCC-ATG). The *Hind*III/*Eco*RI fragment encoding the entire chimaeric protein was subcloned in the mammalian expression vector pcDNA3 (Invitrogen). For cameleon-2nu, the cameleon-2 cDNA was extended by PCR at the 3' end with the sequence encoding the nuclear localization signal (PKKKRKVEDA). Yellow cameleon-3er and -4er cDNA were obtained by extending yellow cameleon-3 cDNA at the 5' end with the sequence encoding the signal peptide from calreticulin (MLLSVPLLLGLGLAAAD), and at the 3' end with the sequence encoding the ER retention signal (KDEL)²⁴. The cDNAs coding for ECFP-calmodulin and M13-EYFP were obtained from yellow cameleon-2 by PCR and cloned into pRSETB behind a polyhistidine tag for bacterial expression and into pcDNA3 behind a Kozak sequence for mammalian expression.

Protein expression, *in vitro* spectroscopy, Ca^{2+} titrations, and reaction kinetics. Chimaeric proteins were expressed in *Escherichia coli*, purified and spectroscopically characterized as previously described². Ca^{2+} titrations were performed by reciprocal dilution of Ca^{2+} -free and Ca^{2+} -saturated buffers¹⁰. *In situ* calibration for $[\text{Ca}^{2+}]$ used the equation²⁹ $[\text{Ca}^{2+}] = K_d'[(R - R_{\min})/(R_{\max} - R)]^{1/n}$, where K_d' is the apparent dissociation constant corresponding to the Ca^{2+} concentration at which R is midway between $R_{\max}$ and $R_{\min}$, and n is the Hill coefficient. Fast kinetic experiments were performed in a stopped-flow spectrofluorometer (Applied Photophysics). After rapid mixing of purified cameleon-1/E104Q (0.8 μM final concentration) and Ca^{2+} of various concentrations in 20 mM MOPS, 100 mM KCl, pH 7.2, at 25 °C, the fluorescence was monitored using an excitation wavelength of 380 nm with an emission cut-on filter of 500 nm. The observed first-order constant (k_{obs}) was calculated from each averaged set of data by nonlinear regression analysis.

Electron microscopy and immunolabelling. Immunogold labelling of

ultrathin sections was performed as described previously³⁰. The polyclonal antibody against GFP was a gift from C. Zuker (University of California, San Diego).

Imaging. Between 2 and 5 days after cDNA transfection with lipofectin (Gibco BRL), HeLa cells at 22 °C were imaged on a Zeiss Axiovert microscope with a cooled CCD camera (Photometrics, Tucson, AZ), controlled by MetaFluor 2.75 software (Universal Imaging, West Chester, PA). Dual-emission ratio imaging of yellow cameleons used a 440DF20 excitation filter, a 455DRLP dichroic mirror, and two emission filters (480DF30 for ECFP, 535DF25 for EYFP) alternated by a filter changer (Lambda 10-2, Sutter Instruments, San Rafael, CA). Interference filters were obtained from Omega Optical and Chroma Technologies (Brattleboro, VT).

Received 8 May; accepted 24 June 1997.

- Heim, R., Prasher, D. C. & Tsien, R. Y. Wavelength mutations and post-translational autooxidation of green fluorescent protein. *Proc. Natl Acad. Sci. USA* **91**, 12501–12504 (1994).
- Heim, R. & Tsien, R. Y. Engineering green fluorescent protein for improved brightness, longer wavelengths and fluorescence energy transfer. *Curr. Biol.* **6**, 178–182 (1996).
- Crivici, A. & Ikura, M. Molecular and structural basis of target recognition by calmodulin. *Annu. Rev. Biophys. Biomol. Struct.* **24**, 85–116 (1995).
- Babu, Y. S., Bugg, C. E. & Cook, W. J. Structure of calmodulin refined at 2.2 Å resolution. *J. Mol. Biol.* **204**, 191–204 (1988).
- Falke, J. J., Drake, S. K., Hazard, A. L. & Peersen, O. B. Molecular tuning of ion binding to calcium signaling proteins. *Q. Rev. Biophys.* **27**, 219–290 (1994).
- Ikura, M. *et al.* Solution structure of a calmodulin-target peptide complex by multidimensional NMR. *Science* **256**, 632–638 (1992).
- Heim, R., Cubitt, A. B. & Tsien, R. Y. Improved green fluorescence. *Nature* **373**, 663–664 (1995).
- Cornack, B. P., Valdivia, R. H. & Falkow, S. FACS-optimized mutants of the green fluorescent protein (GFP). *Gene* **173**, 33–38 (1996).
- Orm6, M. *et al.* Crystal structure of the *Aequorea victoria* green fluorescent protein. *Science* **273**, 1392–1395 (1996).
- Gryniewicz, G., Poonie, M. & Tsien, R. Y. A new generation of Ca^{2+} indicators with greatly improved fluorescence properties. *J. Biol. Chem.* **260**, 3440–3450 (1985).
- Tse, F. W., Tse, A. & Hille, B. Cyclic Ca^{2+} changes in intracellular stores of gonadotropes during gonadotropin-releasing hormone-stimulated Ca^{2+} oscillations. *Proc. Natl Acad. Sci. USA* **91**, 9750–9754 (1994).
- Hofer, A. M. & Schulz, I. Quantification of intraluminal free $[\text{Ca}]$ in the agonist-sensitive internal calcium store using compartmentalized fluorescent indicators: some considerations. *Cell Calcium* **20**, 235–242 (1996).
- Golovina, V. A. & Blaustein, M. P. Spatially and functionally distinct Ca^{2+} stores in sarcoplasmic and endoplasmic reticulum. *Science* **275**, 1643–1648 (1997).
- Montero, M. *et al.* Monitoring dynamic changes in free Ca^{2+} concentration in the endoplasmic reticulum of intact cells. *EMBO J.* **14**, 5467–5475 (1995).
- Kendall, J. M., Badminton, M. N., Sala-Newby, G. B., Campbell, A. K. & Rembold, C. M. Recombinant apoaequorin acting as a pseudo-luciferase reports micromolar changes in the endoplasmic reticulum free Ca^{2+} of intact cells. *Biochem. J.* **318**, 383–387 (1996).
- Porumb, T., Yau, P., Harvey, T. S. & Ikura, M. A calmodulin-target peptide hybrid molecule with unique calcium-binding properties. *Protein Eng.* **7**, 109–115 (1994).
- Maune, J. F., Klee, C. B. & Beckingham, K. Ca^{2+} binding and conformational change in two series of point mutations to the individual Ca^{2+} -binding sites of calmodulin. *J. Biol. Chem.* **267**, 5286–5296 (1992).
- Gao, Z. H. *et al.* Activation of four enzymes by two series of calmodulin mutants with point mutations in individual Ca^{2+} binding sites. *J. Biol. Chem.* **268**, 20096–20104 (1993).
- Martin, S. R. *et al.* Spectroscopic characterization of a high-affinity calmodulin-target peptide hybrid molecule. *Biochemistry* **35**, 3508–3517 (1996).
- Zolotukhin, S., Potter, M., Hauswirth, W., Guy, J. & Muzyczka, N. A “humanized” green fluorescent protein cDNA adapted for high levels of expression in mammalian cells. *J. Virol.* **70**, 4646–4654 (1996).
- Smit, M. J. *et al.* Extracellular ATP elevates cytoplasmic free Ca^{2+} in HeLa cells by the interaction with a 5'-nucleotide receptor. *Eur. J. Pharmacol.* **247**, 223–226 (1993).
- Bootman, M. D., Cheek, T. R., Moreton, R. B., Bennett, D. L. & Berridge, M. J. Smoothly graded Ca^{2+} release from inositol 1,4,5-trisphosphate-sensitive Ca^{2+} stores. *J. Biol. Chem.* **269**, 24783–24791 (1994).
- Brini, M., Marsault, R., Bastianutto, C., Pozzan, T. & Rizzuto, R. Nuclear targeting of aequorin. A new approach for measuring nuclear Ca^{2+} concentration in intact cells. *Cell Calcium* **16**, 259–268 (1994).
- Kendall, J. M., Dormer, R. L. & Campbell, A. K. Targeting aequorin to the endoplasmic reticulum of living cells. *Biochem. Biophys. Res. Commun.* **189**, 1008–1016 (1992).
- Bygrave, F. L. & Benedetti, A. What is the concentration of calcium ions in the endoplasmic reticulum? *Cell Calcium* **19**, 547–551 (1996).
- Button, D. & Eidsath, A. Aequorin targeted to the endoplasmic reticulum reveals heterogeneity in luminal Ca^{2+} concentration and reports agonist- or InS_3 -induced release of Ca^{2+} . *Mol. Biol. Cell* **7**, 419–434 (1996).
- Romoser, V. A., Hinkle, P. M. & Persechini, A. Detection in living cells of Ca^{2+} -dependent changes in the fluorescence emission of an indicator composed of two green fluorescent protein variants linked by a calmodulin-binding sequence. *J. Biol. Chem.* **272**, 13270–13274 (1997).
- Kozak, M. The scanning model for translation: an update. *J. Cell Biol.* **108**, 229–241 (1989).
- Adams, S. R., Bacska, B. J., Taylor, S. S. & Tsien, R. Y. *Fluorescent Probes for Biological Activity of Living Cells—A Practical Guide* (ed. Mason, W. T.) 133–149 (Academic, New York, 1993).
- Balch, W. E., McCaffery, J. M., Plutner, H. & Farquhar, M. G. Vesicular stomatitis virus glycoprotein is sorted and concentrated during export from the endoplasmic reticulum. *Cell* **75**, 841–852 (1994).

Acknowledgements. We thank S. R. Adams and A. B. Cubitt for advice, C. Zuker for the GFP antibody, and C. Klee for the original *Xenopus* calmodulin gene. This work was supported by HHMI (R.Y.T.), NIH (R.Y.T.), HFSP (R.Y.T. and A.M.), MRCC (M.I.) and the Spanish Ministry of Science (J.L.). M.I. is an HHMI international research scholar and MRCC scholar.

Correspondence and requests for materials should be addressed to R.Y.T.

Kinetochores distinguish GTP from GDP forms of the microtubule lattice

Fedor F. Severin*, Peter K. Sorger† & Anthony A. Hyman*

* European Molecular Biology Laboratory, Postfach 10.2209, D-69012 Heidelberg, Germany

† Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

During prometaphase in mitotic cell division, chromosomes attach to the walls of microtubules and subsequently move to microtubule ends, where they stay throughout mitosis^{1,2}. This end-attachment seems to be essential for correct chromosome segregating. However, the mechanism by which kinetochores, the multiprotein complexes that link chromosomes to the microtubules of the mitotic spindle^{3,4}, recognize and stay attached to microtubule ends is not understood. One clue comes from the hydrolysis of GTP that occurs during microtubule polymerization. Although tubulin dimers must contain GTP to polymerize, this GTP is rapidly hydrolysed following the addition of dimers to a growing polymer. This creates a microtubule consisting largely of GDP-tubulin, with a small cap of GTP-tubulin at the end⁵. It is possible that kinetochores distinguish the different structural states of a GTP- versus a GDP-microtubule lattice. We have examined this question *in vitro* using reconstituted kinetochores from the yeast *Saccharomyces cerevisiae*. We found that kinetochores *in vitro* bind preferentially to GTP- rather than GDP-microtubules, and to the plus-end preferentially over the lattice. Our results could explain how kinetochores stay at microtubule ends and thus segregate chromosomes correctly during mitosis *in vivo*. This result demonstrates that proteins exist that can distinguish the GTP conformation of the microtubule lattice.

Chromosome segregation in mitosis is mediated by kinetochores. To monitor the relative attachment of kinetochores to the GDP versus the GTP parts of a microtubule lattice, we constructed segmented microtubules containing regions of both forms of tubulin. As GTP hydrolysis occurs within seconds of microtubule polymerization (the rate of hydrolysis exceeds $4 \times 10^{-2} \text{ s}^{-1}$ (refs 6–8)), we made the GTP-microtubule parts using a slowly hydrolysable analogue, GMPCPP. In previous experiments, the hydrolysis rate of GMPCPP has been measured as $4 \times 10^{-7} \text{ s}^{-1}$ (ref. 9). Microtubules were nucleated from Oregon-green fluorescent GMPCPP tubulin (Fig. 1a, green circles). We measured the rate of hydrolysis under these conditions, confirming that less than 4% of GMPCPP was hydrolysed over 4 h of polymerization (data not shown). GMPCPP microtubules were further elongated by addition of rhodamine GTP tubulin (red circles) which, hydrolysing to GDP after polymerization, creates a GDP segment. To stabilize the GDP regions, microtubules were capped by further addition of GMPCPP tubulin.

To test whether kinetochores can distinguish between GTP- and GDP-tubulin parts of microtubules, we reconstituted kinetochores from *S. cerevisiae in vitro*¹⁰. We began by attaching *S. cerevisiae* centromere DNA to rhodamine-labelled latex beads (Fig. 1a). The beads were incubated in yeast extract, and the bead-bound kinetochore complexes were then mixed with microtubules. Bead attachment to microtubules was monitored by fluorescence microscopy. As previously shown¹⁰, kinetochore complexes assemble on the beads and attach to microtubules.

We monitored the interaction of the reconstituted kinetochores with GMPCPP and GDP parts of microtubule lattice, ignoring any attachment to microtubule ends (see below). A microtubule with a

kinetochore bead bound to a GMPCPP (green) region of the microtubule is shown in Fig. 2a; the GDP-tubulin region and the bead itself are red. After counting the beads bound to GMPCPP or GDP-tubulin and normalizing by the lengths of the microtubules, we found that $92 \pm 4\%$ of the kinetochores bound to the GMPCPP regions of the microtubules (Fig. 2b). Thus an *in vitro* reconstituted kinetochore can distinguish the GTP state from the GDP state of a microtubule lattice ($P < 0.0001$, *t*-test¹¹).

We have shown previously that GMPCPP microtubules have a different structure from GDP microtubules¹². We wondered whether kinetochores were recognizing a unique feature of GMPCPP tubulin or rather a structural motif of the microtubule lattice. Electron microscopy data show that the structure of GDP and GMPCPP microtubules are different¹². Addition of taxol changes GDP microtubules so that their lattice structure resembles that of GMPCPP microtubules¹³. We therefore tested whether kinetochores can bind to a GDP microtubule stabilized with taxol. We made capped microtubules (Fig. 1b) and split them into two portions. We added taxol to one portion and then mixed both portions with kinetochore beads. The percentage of the beads bound to GMPCPP versus GDP tubulin are shown in Fig. 1, and demonstrate that, after taxol was added, kinetochores bind equally to GDP and GMPCPP sections of microtubules ($P > 0.2$, chi-squared test¹¹). Thus it seems that kinetochores recognize a structural motif of microtubules that correlates with microtubule stabilization.

In these experiments we noticed that several beads were bound to the ends of microtubules, but ignored them for the lattice-binding experiments. However, kinetochore microtubules of mitotic spindles have their minus ends at the spindle poles, and their plus ends bound to kinetochores¹⁴. Therefore, a kinetochore might be expected to recognize the plus-ends of microtubules. To quantify the degree to which kinetochores bound to microtubule plus-ends, we polymerized polarity-marked microtubules consisting only of

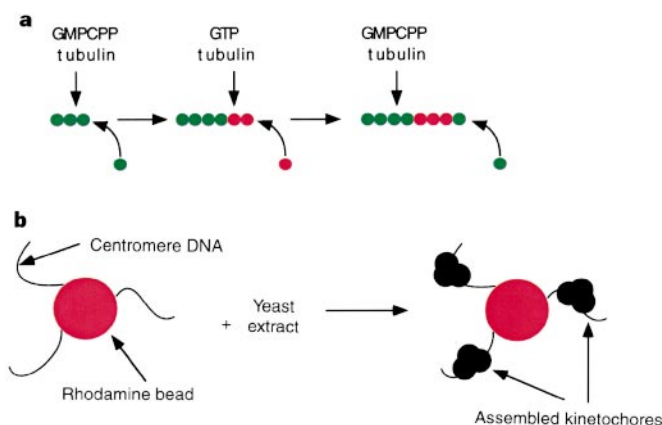


Figure 1 a, Preparation of segmented microtubules. Tubulin labelled with Oregon green was mixed with GMPCPP (green circles). The mixture was incubated at 37°C to allow tubulin to polymerize. The GMPCPP microtubules thus formed were diluted into rhodamine-labelled tubulin (red circles) in the presence of GTP. GTP-tubulin polymerized at the plus-ends of the GMPCPP fragments. GTP-tubulin in polymerized state rapidly hydrolyses the associated GTP molecule into GDP. The GMPCPP-GDP microtubules were diluted into Oregon green-labelled tubulin in the presence of GMPCPP. After polymerization of GMPCPP-tubulin, the microtubules consisted of two GMPCPP fragments at the ends with GDP-tubulin fragments between them. **b**, *In vitro* reconstitution of *S. cerevisiae* kinetochore. The centromere DNA is attached to rhodamine-labelled latex beads by a biotin-streptavidin bond. The beads are incubated in the yeast extract and the kinetochores assemble on the DNA.

GMPCPP tubulin, with a short red minus-end and a long red plus-end, separated by a green segment (Fig. 3a). We mixed kinetochore beads with polarity-marked microtubules and plotted the position of beads along microtubules. The distribution (Fig. 3b) shows that kinetochore beads preferentially bind the microtubule plus-end compared to the rest of the microtubule ($P < 0.01$, t -test¹¹).

Our results show that *in vitro* reconstituted kinetochores from *S. cerevisiae* bind preferentially to GMPCPP microtubules rather than to GDP microtubules, and have a high affinity for microtubule plus-ends. The binding properties of these reconstituted kinetochores are similar to those of isolated mini-chromosomes from yeast¹⁵, and have many of the properties expected of a kinetochore *in vivo*¹⁰. In a different system, a difference in stability of mammalian chromosome-attached microtubule plus-ends compared to minus-ends has been demonstrated³¹, suggesting a difference in binding affinity of kinetochores to microtubule plus-ends, although the issue of whether kinetochores distinguish GTP from GDP lattices was not addressed. Our experiments were performed *in vitro* as such a detailed level of analysis is not possible *in vivo*. *In vivo*, a simple attachment of the microtubule wall matures to a complex end-on attachment in kinetochores of complex eukaryotes, although this yeast has not been seen *in vivo*. Future investigation will determine which aspects of the *in vivo* state have been reconstituted *in vitro*.

The mechanisms by which kinetochores stay at microtubule ends *in vivo* has long been a mystery, but our results suggest that kinetochores could use the structural features of the microtubule lattice to recognize and stay at microtubule ends. This is surprising because most data suggest that there is little GTP at microtubule ends^{6,16}. We can imagine at least two possibilities. First, most data on rates of GTP hydrolysis in microtubules are from *in vitro* studies. It is quite possible that the rate of GTP hydrolysis in microtubules *in vivo* is different or variable, creating a longer GTP cap. The second possibility is that kinetochores bound to microtubule ends could stabilize the GTP state, inhibiting hydrolysis.

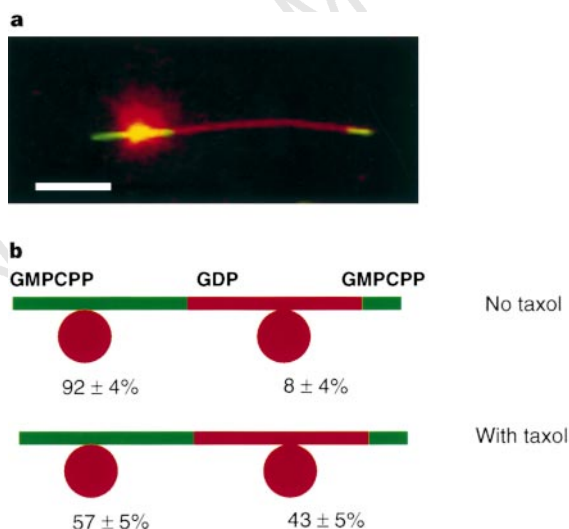


Figure 2 a, Photograph of the capped microtubule with the bead bound to the GMPCPP tubulin region. GMPCPP-tubulin is labelled green, GDP-tubulin and the kinetochore bead are labelled red. Scale bar, 5 μm. **b**, The distribution of kinetochore beads between the regions in the presence and in the absence of taxol. In total we counted 75 beads bound to microtubules, and averages over three experiments were as follows: beads bound to GMPCPP region, 92 ± 4%; beads bound to GDP region, 8 ± 4%. Two more experiments have been done with kinetochore beads binding to capped microtubules in the presence of taxol. In total we counted 64 beads bound, and averages of three experiments were: beads bound to GMPCPP region, 57 ± 5%; beads bound to GDP region, 63 ± 5%.

What are the kinetochore proteins that interact with microtubules? It is known that a complex of four proteins binds to centromere DNA *in vitro* and is required for kinetochore function *in vivo*^{17–20}. It has been shown that the complex is necessary but not sufficient for kinetochores to bind to microtubules *in vitro*^{10,21}, suggesting that other, as-yet unidentified, factors recognize and stabilize the GTP form of microtubules. It is possible that these proteins have homology to the tip attachment complexes in *Tetrahymena* flagella^{22,23}, and one of the *Tetrahymena* proteins also seems to associate with mammalian kinetochores^{22,23}.

In vivo, when kinetochores move away from poles, microtubules polymerize²⁴. We propose that proteins that recognize the GTP form of microtubules provide a powerful mechanism to couple movement and polymerization. If we imagine a kinetochore bound to a microtubule end using such a protein, the kinetochore would be tightly bound (Fig. 4a). Further polymerization would trigger GTP hydrolysis, weakening the interaction (Fig. 4b), allowing the kinetochore to move again until it re-establishes contact with the GTP end (Fig. 4c). Such movement could either be by random diffusion along the microtubule or by microtubule-based motors. We have no evidence yet for plus-end motors on yeast kinetochores, but plus-end motors have been localized to kinetochores in mammalian cells²⁵. Such interactions between motors and proteins that recognize structural features of microtubules could act as a general mechanism by which organelles can be positioned on microtubules.

Our results also have implications for the control of microtubule behaviour. Microtubules turn over by switching constantly between growing and shrinking states^{26,27}, and the rate at which microtubules interconvert between growing and shrinking is known to be regulated, but the mechanisms are obscure^{28,29}. If other microtubule-associated proteins, like the putative kinetochore protein, bind preferentially to GTP tubulin, these proteins could bind to microtubule ends and therefore affect the microtubule

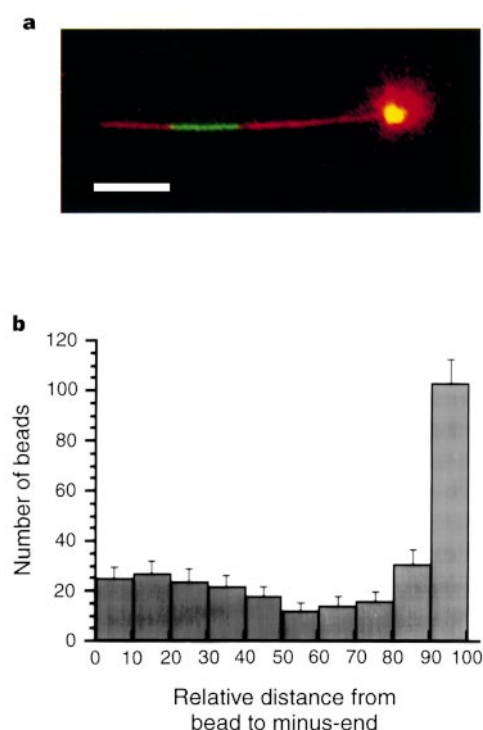


Figure 3 a, The bead bound to the plus-end of the GMPCPP microtubule. Scale bar, 5 μm. **b**, The distribution of the kinetochore beads bound to GMPCPP microtubules. Three independent experiments have been done, and 292 beads counted. The results of the experiments were similar and were added together. Error bars are square roots of the number of beads in each column.

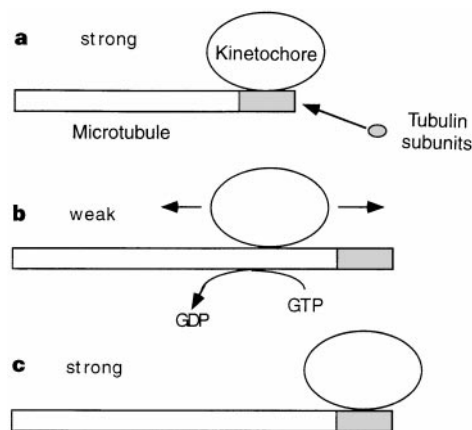


Figure 4 Model for kinetochore behaviour on mitotic-spindle microtubule. **a**, The kinetochore binds to the GTP part of the microtubule. **b**, The microtubule polymerizes by end-on addition of GTP-tubulin subunits and GTP hydrolyses, weakening the attachment of the kinetochore to the microtubule. **c**, The kinetochore moves to the end of the microtubule, where it reattaches to GTP-tubulin.

dynamics at substoichiometric amounts. Although most microtubule-associated proteins so far isolated bind along microtubule lattices, it is important to remember that all such preparations have been made using taxol-stabilized microtubules. Because we have shown that taxol mimics the GTP state of microtubules, it remains possible that known microtubule-associated proteins can distinguish between GTP and GDP states of microtubules. □

Methods

In vitro reconstitution of *S. cerevisiae* kinetochores. Yeast extracts were made as described¹⁰ except that yeast cells were broken in a buffer containing (in mM) 50 bis-Tris-propane, pH 7.0, 100 β -glycerophosphate, 2 EDTA, 2 EGTA, 200 KCl 10% glycerol. *CEN3* DNA was attached to rhodamine latex beads as described¹⁰. To assemble the kinetochore complexes on the beads, 200 μ g of extract was added to 10^7 beads and 2 μ g of sonicated salmon sperm DNA in kinetochore buffer (10 mM HEPES, pH 8.0, 6 mM $MgCl_2$, 10% glycerol) and incubated for 30 min at room temperature.

Preparation of capped microtubules. Tubulin was prepared and labelled with rhodamine or Oregon green (Molecular Probes, Eugene) as described³⁰. 1 μ M Oregon green tubulin was added to 0.5 mM GMPCPP in BRB80 (80 mM potassium PIPES, pH 6.8, 1 mM $MgCl_2$, 1 mM EGTA) at 37°C. 1 μ M Oregon green tubulin was added to the mixture every 30 min, for 2–3 h, creating GMPCPP microtubules with an average length of 6.5 μ m. Then 2.0 μ l of 0.1 mM GTP, 10 μ M tubulin, 5 μ M rhodamine tubulin in BRB80 were preincubated for 10 min on ice and pre-warmed for 30 s at 37°C. 2 μ l Oregon green microtubules was added to the mixture and GTP tubulin was allowed to polymerize for 20 min. Under these conditions, 94% ($n = 69$) of the GMPCPP microtubules nucleated GTP microtubules. 40 μ l of 0.5 μ M Oregon green tubulin, 0.5 mM GMPCPP in BRB80 were preincubated on ice for 10 min, pre-warmed at 37°C for 1 min and mixed with 0.5 μ l of the GMPCPP–GTP microtubules, giving a final GTP concentration of 1 μ M and a GMPCPP concentration of 500 μ M. The mixture was replenished with 0.5 μ M Oregon green tubulin after 1 h of incubation at 37°C and left at 37°C for 30 min. For the taxol experiment, 20 μ l of microtubules was mixed with either 20 μ l of BRB80 or 20 μ l of 50 μ M taxol in BRB80. Taxol microtubules were centrifuged for 5 min at 50,000 r.p.m. in a TLA 100.1 rotor and resuspended in 15 μ l of BRB80 with or without 10 μ M taxol.

Preparation of GMPCPP microtubules. 1 μ M Oregon green tubulin was added to 0.5 mM GMPCPP in BRB80 (80 mM potassium PIPES, pH 6.8, 1 mM $MgCl_2$, 1 mM EGTA) at 37°C. 1 μ M Oregon green tubulin was added to the mixture every 30 min for 3–4 h. Then 2 μ l of the Oregon green microtubules was added to 20 μ l of 0.3 μ M tubulin, 0.2 μ M rhodamine tubulin, 0.5 mM

GMPCPP in BRB80. The mixture was incubated at 37°C for 30 min and replenished with 0.3 μ M of unlabelled tubulin and 0.2 μ M of rhodamine tubulin every 30 min for 3–4 h. As tubulin grows approximately 3 times faster from the plus- compared to the minus-end²⁷, the red fragments at the minus-ends of the microtubules were visibly shorter than those at the plus-ends. The polarity of microtubules was confirmed in a kinesin gliding assay.

Microtubule binding assays. 2 μ l of kinetochore beads was mixed with 2 μ l of capped microtubules. The mixture was incubated for 5 min at room temperature, diluted with 40 μ l of BRB80 and observed with a Zeiss Axioskop microscope and a 63 $\times$ PlanApo 1.6 lens using a Colour Cool View camera (Photonic Sciences, East Sussex). To avoid bias, all microtubules containing a bead were photographed. To examine the microtubule colour to which the beads were bound, the colours were split into the three components of the red–green–blue (RGB) signal. Alternatively, 1 μ l of the GMPCPP microtubules was mixed with 10 μ l of the kinetochore beads, allowed to bind for 5 min, then diluted with 200 μ l of BRB80 and observed the same way.

Data analysis. For the experiments investigating binding to GMPCPP versus GDP lattice, we counted only the beads that were visibly not bound to the end, so beads that obscured the microtubule end when bound were not counted. To normalize the bead numbers bound to either type of the lattice, the total numbers of beads bound to GMPCPP and GDP fragments were multiplied by the total lengths of GMPCPP and GDP fragments, so the percentage of beads bound to the GMPCPP fragments was $100 \times N_c L_d / (N_c L_d + N_d L_c)$, and the percentage of the beads bound to the GDP fragments was $100 \times N_d L_c / (N_c L_d + N_d L_c)$, where N_c is 67 (the number of beads bound to GMPCPP fragments), N_d is 8 (the number of beads bound to GDP fragments), L_c is 485 μ m (the total length of the GMPCPP fragments), and L_d is 539 μ m (the total length of the GDP fragments). Three independent experiments were done, and data added together and analysed. For the experiments on the effect of taxol addition, 62 beads were counted. To analyse the binding of beads to the plus-end, the minus-end and the lattice, we used GMPCPP microtubules. A microtubule was counted as binding to the end if it was obscuring the end of the microtubule. Only microtubules that had a clearly distinguishable polarity were included ($\sim 90\%$). Three independent experiments were done, and all data are summarized in Fig. 3b, for a total of 292 beads. The distance from the microtubule minus-end to the centre of the bead bound was divided by the length of the microtubule and expressed as a percentage.

Received 11 April; accepted 30 May 1997.

- Rieder, C. L. & Alexander, S. P. Kinetochores are transported poleward along a single astral microtubule during chromosome attachment to the spindle in newt lung cells. *J. Cell Biol.* **110**, 81–95 (1990).
- Skibbens, R. V., Skeen, V. P. & Salmon, E. D. Directional instability of kinetochore motility during chromosome congression and segregation in mitotic newt lung cells: A push–pull mechanism. *J. Cell Biol.* **122**, 859–875 (1993).
- Earnshaw, W. C. & Tomkiel, J. E. Centromere and kinetochore structure. *Curr. Biol.* **4**, 86–93 (1992).
- Bloom, K. The centromere frontier: kinetochore components, microtubule-based motility, and the CEN-value paradox. *Cell* **73**, 621–624 (1993).
- Kirschner, M. W. & Mitchison, T. J. Beyond self assembly: from microtubules to morphogenesis. *Cell* **45**, 329–342 (1986).
- Schilstra, M. J., Martin, S. R. & Bayley, P. M. On the relationship between nucleotide hydrolysis and microtubule assembly: studies with a GTP-regenerating system. *Biochem. Biophys. Res. Comm.* **147**, 588–595 (1987).
- O'Brien, E. T., Voter, W. A. & Erickson, H. P. GTP hydrolysis during microtubule assembly. *Biochemistry* **26**, 4148–4156 (1987).
- Stewart, R. J., Farrell, K. W. & Wilson, L. Role of GTP hydrolysis in microtubule polymerisation: evidence for a coupled hydrolysis mechanism. *Biochemistry* **29**, 6489–6498 (1990).
- Hyman, A. A., Salser, S., Drechsel, D. N., Unwin, N. & Mitchison, T. J. Role of GTP hydrolysis in microtubule dynamics: information from a slowly hydrolyzable analogue, GMPCPP. *Mol. Biol. Cell.* **3**, 1155–1167 (1992).
- Sorger, P. K., Severin, F. F. & Hyman, A. A. Factors required for the binding of reassembled yeast kinetochores to microtubules *in vitro*. *J. Cell Biol.* **127**, 995–1008 (1994).
- Bailey, N. *Statistical methods in Biology* (Cambridge Univ. Press, 1995).
- Hyman, A. A., Chretien, D., Arnal, I. & Wade, R. H. Structural changes accompanying GTP hydrolysis in microtubules: information from a slowly hydrolyzable analogue guanylyl-(α , β)-methylene-diphosphonate. *J. Cell Biol.* **128**, 117–125 (1995).
- Arnal, I. & Wade, R. H. How does taxol stabilize microtubules? *Curr. Biol.* **5**, 900–908 (1995).
- McIntosh, J. R. & Euteneuer, U. Tubulin hooks as probes for microtubule polarity: an analysis of the method and an evaluation of data on microtubule polarity in the mitotic spindle. *J. Cell Biol.* **98**, 525–533 (1984).
- Kingsbury, J. & Koshland, D. Centromere-dependent binding of yeast minichromosomes to microtubules *in vitro*. *Cell* **66**, 483–495 (1991).
- Mitchison, T. J. & Kirschner, M. W. Dynamic instability of microtubule growth. *Nature* **312**, 237–242 (1984).
- Lechner, J. & Carbon, J. A 240 kD multisubunit protein complex, CBF3, is a major component of the budding yeast centromere. *Cell* **64**, 717–726 (1991).
- Jiang, W., Lechner, J. & Carbon, J. Isolation and characterization of a gene (CBF2) specifying a protein component of the budding yeast centromere. *J. Cell Biol.* **121**, 513–519 (1993).

19. Strunnikov, A. V., Kingsbury, J. & Koshland, D. CEP3 encodes a centromere protein of *Saccharomyces cerevisiae*. *J. Cell Biol.* **128**, 749–760 (1995).
20. Connelly, C. & Heiter, P. Budding yeast SKP1 encodes an evolutionary conserved protein required for cell cycle progression. *Cell* **86**, 275–285 (1996).
21. Kingsbury, J. & Koshland, D. Centromere function on minichromosomes isolated from budding yeast. *Mol. Biol. Cell* **4**, 859–870 (1993).
22. Miller, J. M., Wang, W., Balczon, R. & Dentler, W. L. Ciliary microtubule capping structures contain a mammalian kinetochore antigen. *J. Cell Biol.* **110**, 703–714 (1990).
23. Wang, W., Suprenant, K. & Dentler, W. Reversible association of a 97-kDa protein complex found at the tips of ciliary microtubules with *in vitro* assembled microtubules. *J. Biol. Chem.* **268**, 24796–24807 (1993).
24. Mitchison, T., Evans, L., Schulze, E. & Kirschner, M. Sites of microtubule assembly and disassembly in the mitotic spindle. *Cell* **45**, 515–527 (1986).
25. Hyman, A. A. & Mitchison, T. J. Two different microtubule-based motor activities with opposite polarities in kinetochores. *Nature* **351**, 206–211 (1991).
26. Cassimeris, L., Pryer, N. K. & Salmon, E. D. Real-time observations of microtubule dynamic instability in living cells. *J. Cell Biol.* **107**, 2223–2231 (1988).
27. Walker, R. A. *et al.* Dynamic instability of individual microtubules analysed by video light microscopy: rate constants and transition frequencies. *J. Cell Biol.* **107**, 1437–1448 (1988).
28. Belmont, L. D., Hyman, A. A., Sawin, K. E. & Mitchison, T. J. Real-time visualization of cell cycle-dependent changes in microtubule dynamics in cytoplasmic extracts. *Cell* **62**, 579–589 (1990).
29. Sammak, P. J. & Borisy, G. G. Direct observation of microtubule dynamics in living cells. *Nature* **332**, 724–726 (1988).
30. Hyman, A. A. *et al.* Preparation of modified tubulins. *Methods Enzymol.* **196**, 478–485 (1991).
31. Huitorel, P. & Kirschner, M. W. The polarity and stability of microtubule capture by the kinetochore. *J. Cell Biol.* **106**, 151–160 (1988).

Acknowledgements. We thank S. Schuyler for early experiments suggesting the role of lattice structure; J. Howard for help with statistics; and M. Glotzer, P. Gonczy, S. Eaton, K. Simons, R. Heald, S. Reinsch and S. Andersen for critical reading of the manuscript.

Correspondence and requests for materials should be addressed to A.A.H.

Mutant analysis links the translocon and BiP to retrograde protein transport for ER degradation

Richard K. Plemper*, Sigrun Böhmler*, Javier Bordallo*, Thomas Sommer† & Dieter H. Wörz

* Institut für Biochemie der Universität Stuttgart, Pfaffenwaldring 55, D-70569 Stuttgart, Germany

† Max-Delbrück-Centrum für Molekulare Medizin, Robert-Rössle Strasse 10, D-13122 Berlin, Germany

Proteins enter the secretory pathway through the endoplasmic reticulum¹, which delivers properly folded proteins to their site of action² and contains a quality-control system to monitor and prevent abnormal proteins from being delivered³. Many of these proteins are degraded by the cytoplasmic proteasome^{4–8}, which requires their retrograde transport to the cytoplasm^{5,6}. Based on a co-immunoprecipitation of major histocompatibility complex (MHC) class I heavy-chain breakdown intermediates with the translocon subunit Sec61p (refs 9, 10), it was speculated that Sec61p may be involved in retrograde transport¹¹. Here we present functional evidence from genetic studies that Sec61p mediates retrograde transport of a mutated luminal yeast carboxypeptidase ycsY (CPY*) *in vivo*. The endoplasmic reticulum luminal chaperone BiP (Kar2p) and Sec63p, which are also subunits of the import machinery^{10,12}, are involved in export of CPY* to the cytosol. Thus our results demonstrate that retrograde transport of proteins is mediated by a functional translocon. We consider the export of endoplasmic reticulum-localized proteins to the cytosol by the translocon for proteasome degradation to be a general process in eukaryotic cell biology.

Proteins cross the membrane of the endoplasmic reticulum (ER) either co-translationally or post-translationally^{9,10}. For both translocation routes, the Sec61 heterotrimeric complex consisting of yeast proteins Sec61p, Sbh1p and Sss1p (ref. 10) constitutes the membrane channel. The post-translational pathway depends on an additional tetrameric complex consisting of the ER transmembrane proteins Sec62p, Sec63p and Sec71p, together with the peripheral membrane protein Sec72p (refs 10, 12). Sec62p, Sec71p and Sec72p

provide a cytosolic binding site for the newly synthesized precursor molecules, whereas Sec63p and the luminal chaperone Kar2p (BiP) mediate the driving force for import^{10,12}.

We analysed retrograde transport using yeast strains carrying mutant alleles of translocon components, and used mutated yeast

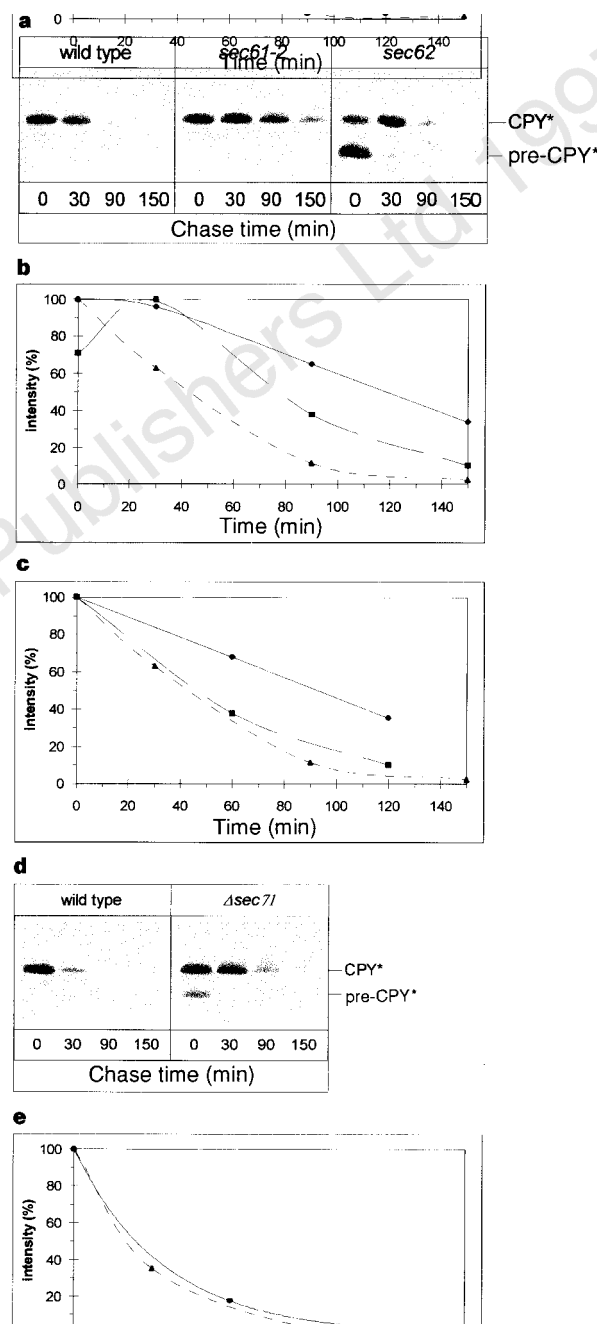


Figure 1 A functional Sec61p is required for ER degradation. **a–c**, CPY* is stabilized in the *sec61-2* mutant, pre-CPY* import and not degradation is affected in the *sec62* mutant. Pulse-chase analysis of CPY* was performed using the isogenic strains W303-1C (wild type, triangles), YRP086 (*sec61-2*, diamonds) and YRP087 (*sec62*, squares). Cells were grown at the permissive temperature of 25°C. After the indicated chase times the cells were lysed, CPY* was immunoprecipitated, and the antigenic material was separated by SDS-PAGE and analysed using a Molecular Dynamics imaging system. **d, e**, A *sec71* knockout leads also to accumulation of pre-CPY* but not to a defect in CPY* degradation. Pulse-chase analysis of the isogenic strains W303-1C (wild type, triangles) and YRP091 ($\Delta sec71$, circles) was done at 30°C as described in **a**.

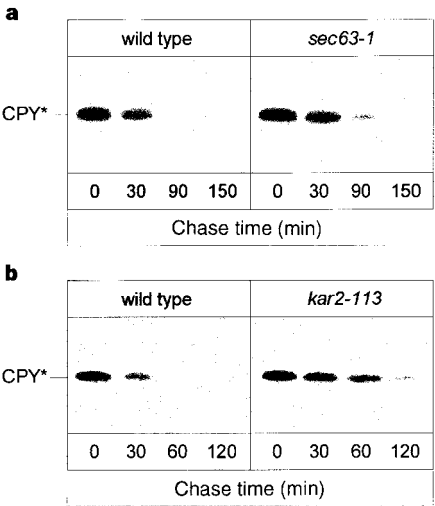


Figure 2 ER degradation of CPY* is affected by the *sec63-1* (**a**) and *kar2-113* (**b**) mutations. Pulse-chase analysis of the isogenic strains W303-1C (wild type), YRP088 (*sec63-1*) and YRP146 (*kar2-113*) at the permissive temperature of 25 °C was performed.

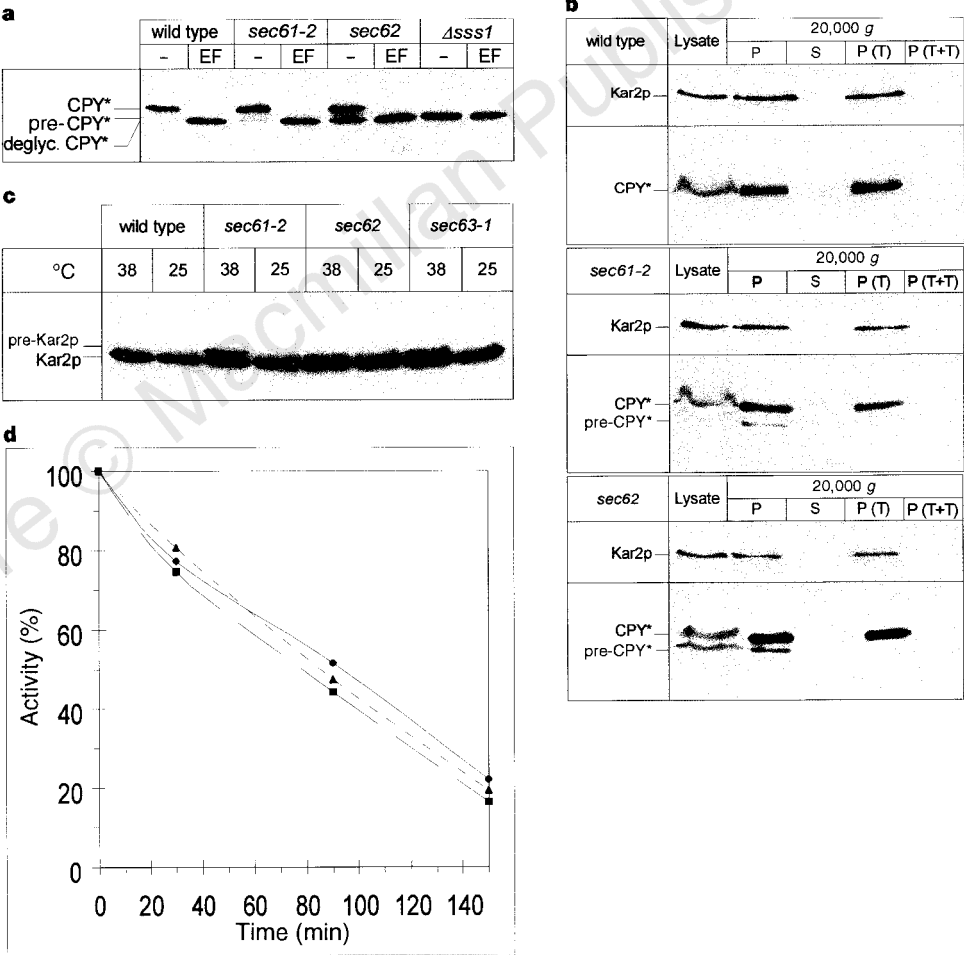


Figure 3 Control experiments demonstrate that CPY* stabilization in the *sec61-2* background is based on a defective export. **a**, CPY* accumulating in *sec61-2* mutant cells is glycosylated, whereas the second antigenic species is unglycosylated, showing that it is pre-CPY*. Western analysis of CPY* and pre-CPY* in the isogenic strains W303-1C (wild type), YRP086 (*sec61-2*), YRP087 (*sec62*) and YRP090 (Δ *sss1*, pGALSSS1 on plasmid) after growth of cells at 32 °C. Deglycosylation was performed with endoglycosidase F (EF) after immunoprecipitation of CPY* antigenic material. **b**, Although CPY* is membrane protected from proteolysis, pre-CPY* is bound to the cytosolic face of vesicles. Western analysis of CPY*, pre-CPY* and (for control) Kar2p, in the strains W303-1C (wild type), YRP086 (*sec61-2*) and YRP087 (*sec62*), after preparing spheroblast homo-

genes followed by a 20,000g centrifugation. (P, pellet fraction; S, supernatant; P (T), pellet treated with trypsin; P (T + T), pellet treated with trypsin and Triton X-100. **c**, Kar2p is imported without precursor accumulation in *sec61-2* mutant cells at the permissive temperature. Western analysis of Kar2p in the strains W303-1C (wild type), YRP086 (*sec61-2*), YRP087 (*sec62*) and YRP088 (*sec63-1*). **d**, The Ubc6-Ubc7 proteasome degradation system is functional in the *sec61-2* and *sec63-1* mutants. β -Galactosidase activity was tested after alkaline lysis of the strains W303-1C (wild type, triangles), YRP086 (*sec61-2*, diamonds) and YRP088 (*sec63-1*, squares). Cells expressing the plasmid-encoded fusion protein Deg1- β -galactosidase were analysed after incubation at the permissive temperature of 25 °C.

carboxypeptidase yscY (CPY*) as a substrate for ER degradation¹³. When following the fate of CPY* in a strain harbouring the temperature-sensitive *sec61-2* allele¹⁴ by pulse-chase analysis at 25 °C, a considerable stabilization of CPY* compared to the wild type was visible (Fig. 1a). The half-life of CPY* increased about threefold in the *sec61-2* background (Fig. 1b), indicating a disturbed delivery of luminal CPY* to the cytosolic proteasome. In contrast, at the same temperature of 25 °C, a mutation in Sec62p led to an import defect, clearly visible by the accumulation of pre-CPY* (Fig. 1a) and the appearance of an intensity maximum of luminal CPY* after 30 min chase time (Fig. 1b). A delayed import was also seen in a Sec71p deletion strain (Fig. 1d), resulting in an apparent stabilization of CPY*. However, in the *sec62* (Fig. 1c) and Δ *sec71* (Fig. 1e) mutants, the kinetics of CPY* degradation was essentially identical to wild type. This was demonstrated by taking the second chase point (30 min), at which the precursor of CPY* is completely imported into the ER, as the time point from which the half-life was calculated. Applying the same calculation to the *sec61-2* mutant strain, the considerably reduced degradation rate of CPY* compared with wild type was again apparent (Fig. 1c). As the function of Sec72p depends on the binding to Sec71p, the *sec71* deletion strain is also devoid of Sec72p function. Thus Sec72p cannot be involved in CPY* retrograde transport. Also, deletion of *SBH1*, a non-essential subunit of the Sec61 heterotrimeric complex^{15,16}, did not show any changes in the steady-state level of CPY* (data not shown). Furthermore, the amount of CPY* was unaffected in a strain carrying a deletion in *SSH1*, a non-essential homologue of *SEC61*, which appears to be capable of mediating exclusively co-translational protein translocation¹⁶, as well as in a strain deleted in *SBH2*, which encodes a homologue of *SBH1* (ref. 16, and data not shown).

Sec63p and the chaperone Kar2p regulate the dynamics of the Sec61p channel¹². Interaction between both proteins occurs through a DnaJ box located within the luminal tail of Sec63p (refs 17, 18), a common binding motif for members of the Hsp70 family¹⁹. Because the average diameter of the Sec61p channel is about 20 Å (ref. 20),

export of CPY* in a globular state seems unlikely. Kar2p, bound to Sec63p, might unfold CPY* before export. We thus tested the degradation rate of CPY* in a *sec63-1* background carrying a mutation in the DnaJ box which prevents binding of Kar2p²¹ by pulse-chase analysis at the permissive temperature of 25 °C. Stabilization of CPY* was observed in the *sec63-1* mutant (Fig. 2a), but no CPY* precursor accumulated. Compared to the wild type, the half-life increased about 1.5-fold, indicating that *sec63p* functions in the CPY* export. We directly addressed the question of whether the presence of intact Kar2p is required for retrograde transport by examining the influence of the mutated *kar2-113* allele²² on the degradation rate of CPY*. Although import of CPY* into the ER was unaffected in *kar2-113* mutant cells at 25 °C, a remarkable stabilization of luminal CPY* of about twofold was observed (Fig. 2b). This effect may be due to a defective regulatory function of the Kar2-113 protein on Sec61p, or a reduced unfolding capacity on CPY*.

It was important to demonstrate that the molecular form of CPY* seen in the *sec61-2* mutant represents glycosylated, luminal CPY*, and that the additional molecular form seen in the *sec62* mutant was cytosolic pre-CPY* and not some already exported degradation intermediate. A CPY* species, once imported into the ER lumen, is expected to be glycosylated and thus accessible to deglycosylation by endoglycosidase F, whereas non-imported pre-CPY* should be devoid of carbohydrate. We induced the accumulation of CPY* precursor molecules in the *sec61-2* and the *sec62* mutant by incubating cells at 32 °C. In *sec63-1* and *kar2-113* mutant cells, no pre-CPY* was visible at this temperature (data not shown). The antigenic material of higher relative molecular mass (M_r) appearing in all strains is completely accessible to endoglycosidase F treatment (Fig. 3a). In contrast, the band of lower M_r seen in the *sec61-2* and *sec62* mutants does not undergo a shift upon endoglycosidase F treatment as expected for non-imported pre-CPY*. We confirmed the stability of this band against deglycosylation by analysing the CPY* species in a strain devoid of Sss1p, an essential subunit of the translocon²³. This strain was kept alive with the *SSS1* gene under the control of the GAL promoter. After repressing the promoter by adding glucose, the protein import into the ER was completely blocked. In these cells only pre-CPY* accumulated, which, as expected, was completely resistant to endoglycosidase F treatment (Fig. 3a).

We then determined the intracellular localization of the CPY* species in *sec61-2*, *sec62* and in wild-type cells at 32 °C. After destruction of cells, the lysate was separated into a pellet fraction containing the microsomes and a supernatant fraction. The fractions were run on an SDS-polyacrylamide gel and immunoblotted with specific antibodies against CPY* and Kar2p as a control for the integrity of the ER vesicles. The *sec61-2* and *sec62* mutants accumulate pre-CPY* and CPY* species, which fractionate with the pellet (Fig. 3b). Trypsin treatment of the pellet leads to disappearance of the predicted pre-CPY*, whereas the band expected to represent luminal CPY* remained stable. Only after treatment with Triton X-100 was the latter species accessible to trypsin digestion. This indicates that pre-CPY* was attached to the cytosolic face of the ER membrane, probably to the Sec62–Sec71–Sec72 complex¹²; CPY* was imported into the ER lumen in both mutant strains.

Investigation of the import of two additional proteins, Kar2p wild-type protein and proteinase yscA, showed that the decreased degradation rate of CPY* seen in the strains carrying the mutated translocon subunits *sec61-2* and *sec63-1* was not due to some general import defect that might affect components of the ER degradation system. Although, as expected, pre-Kar2p accumulated at the restrictive temperature of 38 °C in the mutant cells (Fig. 3c), no accumulation was seen at the permissive temperature (25 °C), as had been found for CPY* (Figs 1a, 2a). There was also no accumulation of prepro-proteinase yscA in the mutants at 25 °C (data not shown). In *sec62* mutant cells, only a very weak accumulation of pre-Kar2p could be observed at 25 °C, suggesting that Kar2p was mainly

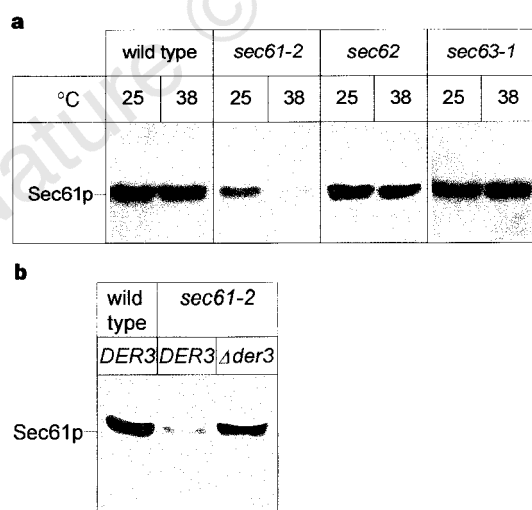


Figure 4 Mutant Sec61p is a substrate of ER degradation. **a**, The Sec61-2p steady-state level is reduced at the permissive temperature of 25 °C. Western analysis of Sec61p in the strains W303-1C (wild-type), YRP086 (*sec61-2*), YRP087 (*sec62*) and YRP088 (*sec63-1*) at permissive (25 °C) and restrictive (38 °C) temperature. Cells were grown at 25 °C and portions were shifted to the restrictive temperature. After 60 min, crude extracts of cells incubated at 25 °C or shifted to 38 °C were analysed. **b**, Der3, a gene product involved in CPY* degradation, participates in the degradation of mutant Sec61p. Western analysis of Sec61p in the isogenic strains W303-1C (wild-type), YRP086 (*sec61-2*) and YRP105 (Δ *der3 sec61-2*). Cells were analysed after a shift to the restrictive temperature of 38 °C for 45 min.

imported through the co-translational translocation pathway²⁴. These findings suggest that the import machinery into the ER lumen in the *sec61-2* and the *sec63-1* mutants was unimpaired at the permissive temperature of 25 °C.

The ubiquitin-conjugating enzymes Ubc6 and Ubc7 act at the ER membrane. Together with the proteasome they are essential components of the CPY* degradation process⁵. We had therefore to exclude the possibility that this degradation system was impaired by the mutated transmembrane proteins *sec61-2* or *sec63-1*. Another protein exclusively degraded through the Ubc6, Ubc7 proteasome system is a fusion between the Deg1 domain of the MAT α 2 transcriptional repressor and β -galactosidase²⁵. We therefore tested degradation of this fusion protein in *sec61-2*, *sec63-1* and in wild-type cells at 25 °C. The decrease of β -galactosidase activity was equal in all strains (Fig. 3d). Also, when analysed in a pulse-chase experiment, the fusion protein was degraded in both mutants as in wild-type cells (data not shown). These experiments demonstrate that the proteolytic machinery was intact in the *sec61-2* and *sec63-1* strains. This result provides further support for the idea that the stabilization of luminal CPY* in the *sec61-2* and the *sec63-1* background is due to a delayed export of the protein into the cytosol.

The *sec61-2* allele codes for a mutated Sec61 protein, which itself undergoes ER-associated degradation at the restrictive temperature of 38 °C (ref. 4). We examined the Sec61p steady-state level at this temperature and at 25 °C, at which degradation of CPY* is already dramatically retarded (Fig. 1a). Even at 25 °C, the steady-state level of Sec61p was considerably reduced as compared with the wild type, but no reduction of Sec61p was seen in *sec62* or *sec63-1* mutant cells (Fig. 4a). Therefore, even at the permissive temperature of 25 °C, proteasome-dependent degradation of the mutated Sec61 protein seems to occur. We cannot completely exclude the possibility that the *sec61-2* mutation directly affects the binding of a protein required for the export process, but the finding of a lowered steady-state level of Sec61p in the *sec61-2* mutant cells might provide a more likely explanation for our results: that the retrograde transport of CPY* is affected at 25 °C in the mutant background, whereas import is not. In wild-type cells, about 40–60% of the total Sec61p population was found to be either associated with ribosomes or assembled with the Sec62p/Sec63p subcomplex. About 30–50% of Sec61p remains assembled in the trimeric complex, but is probably not involved in protein translocation¹⁵. This latter portion of Sec61p may be recruited easily for the export of ER degradation substrates without affecting import in wild-type cells. However, in the *sec61-2* mutant at 25 °C, the free pool of Sec61p may be reduced, but the absolute amount of Sec61p transloci involved in co- or post-translational transport is not affected, as protein translocation is essential for cell viability. In contrast, ER-associated degradation is not essential for viability^{4,5,26}. Because priority is given to protein import, the cell must have tools to tightly regulate import as well as export. Taking this model into account, we cannot completely exclude the possibility that the export defects of CPY* seen with the *sec63-1* and *kar2-113* mutants also have their basis in a reduced number of transloci available for retrograde transport. However, given that *sec63-1* and *kar2-113* mutant cells do not accumulate CPY* precursor molecules even at 30 °C (data not shown), we assume that there is a large amount of free transloci in these mutants at 25 °C, the temperature at which the experiments were done. We therefore suppose that Kar2p and Sec63p are directly involved in the retrograde transport of CPY*.

Like the degradation of CPY* (ref. 5), proteolysis of unassembled Sec61p occurs through the Ubc6–Ubc7 proteasome degradation pathway⁴. We investigated whether export and degradation of Sec61-2p also depends on additional components known to be required for CPY* degradation. The *DER3* gene product, which is identical to that of *HRD1* (ref. 27), has been shown to reside in the ER membrane, and is essential for degradation of CPY* (J.B., R.K.P., A. Finger and D.H.W., unpublished data). We introduced a null

allele of *DER3* into the *sec61-2* mutant strain. Although at the restrictive temperature of 38 °C the steady-state level of Sec61-2p is dramatically lowered in *DER3* cells, Sec61-2p reaches wild-type levels in the *der3* deletion strain (Fig. 4b). Together with the finding that, in the absence of Der3p, *sec61-2* cells are able to grow at 38 °C (J.B., R.K.P., A. Finger and D.H.W., unpublished data), these results indicate that Der3p is also required for the degradation of mutated Sec61p. This finding demonstrates that an intact ER-degradation machinery is present in *sec61-2* mutant cells, even at 38 °C. Indeed, Der3p is present in the *sec61-2* background in wild-type amounts (data not shown). It seems likely that under restrictive conditions, and to a lesser extent under the permissive conditions of 25 °C, unassembled Sec61p chains enter, probably through lateral gating, still intact transloci to become accessible to the proteasome.

It has also been proposed that retrograde transport requires Sec61p in higher eukaryotic cells¹¹. In cells expressing the human cytomegalovirus protein US2, or treated with dithiothreitol, the MHC class I heavy chain is rapidly degraded. When proteasome inhibitors were applied, a non-glycosylated cytosolic breakdown intermediate accumulated in US2-expressing cells. This intermediate was co-immunoprecipitated with Sec61p. In cells treated with dithiothreitol, full-length glycosylated heavy chain was found to be complexed with Sec61p. Our results provide functional support for the idea that components of the translocon are involved in retrograde protein translocation. Although we cannot exclude categorically some indirect effects, our data strongly suggest that Sec61p, and probably also Sec63p and Kar2p, are components of a universally acting subcomplex mediating retrograde transport out of the ER for subsequent degradation of proteins that are not properly folded. It is possible that the direction of transport through the ER membrane is determined by the subunit composition of the translocon. Several transmembrane proteins involved in ER degradation have been described, but their detailed function is still unknown^{26,27}. These proteins could interact with the central components of the Sec61p translocon, resulting in a complex that is then programmed for retrograde transport. The nature of this complex may be probed once it becomes possible to isolate transloci in the state of export. □

Methods

Construction of strains and plasmids. Genetic experiments were performed using standard methods²⁸. The wild-type strain used in this study was W303-1C (*MAT α* , *ade2-loc ura3-1 his3-11,15 leu2-3,112 trp1-1 can1-100 prc1-1*)²⁶. W303-1Ca (*MAT α* , *ade2-loc ura3-1 his3-11,15 leu2-3,112 trp1-1 can1-100 prc1-1*) was provided by M. M. Hiller. YFP338 (*MAT α* , *sec61-2 leu2-3,112 ura3-52 trp1-1 his4-401 HOL1-1*) was provided by M. D. Rose. RSY529 (*MAT α* , *sec62 leu2-3,112 his4-401 ura3-52*), RSY151 (*MAT α* , *sec63-1 leu2-3,112 ura3-52 pep4-3*) and RSY926 (*MAT α* , Δ *sec71::LEU2 ura3-52 lys2-801 his3 Δ 200 leu2 Δ 1 trp1 Δ 63 ade2-101*) were provided by R. Schekman. YRP086 (*MAT α* , *sec61-2 ade2-loc ura3his3-11,15 leu2-3,112 trp1-1 can1-100 prc1-1*), YRP087 (*MAT α* , *sec62 ade2-loc ura3 his3-11 leu2-3,112 trp1-1 can1-100 prc1-1*), YRP088 (*MAT α* , *sec63-1 ade2-loc ura3 his3-11,15 leu2-3,112 trp1-1 can1-100 prc1-1*) and YRP090 (*MAT α* , Δ *ss1::ADE2 pGALSSS1 (LEU2 ARS/CEN (pKF31)) ade2-loc ura3-1 his3-11,15 leu2-3,112 trp1-1 can1-100 prc1-1*) were derived from 4-fold crosses of YFP338, RSY529, RSY151 and YTX138 (Δ *ss1::ADE2 (LEU2 ARS/CEN (pKF31)) his3-11-15 trp1-1 ura3-1ade2-1 can1-100*) with W303-1C and W303-1Ca, respectively, followed by selection for the temperature-sensitive phenotype and the W303-1C markers. The plasmid pKF31 was a gift from K. Finke. YRP091 (*MAT α* , Δ *sec71::LEU2 ade2-loc ura3-1 his3-11,15 leu2-3,112 trp1-1 can1-100 prc1-1*) was constructed by transforming W303-1C with the *sec71::LEU2* construct isolated from RSY926 by PCR. YRP146 (*MAT α* , *kar2-113 ade2-loc ura3-1 his3-11,15 leu2-3,112 trp1-1 can1-100 prc1-1*) was constructed as described²⁹, and plasmid MR780 was provided by M. D. Rose. YRP105 (*MAT α* , Δ *der3::HIS3 sec61-2 ade2-loc ura3-1 his3-11,15 leu2-3,112 trp1-1 can1-100 prc1-1*) was constructed by transforming YRP086 with plasmid puc-del3 (*der3::HIS3*). That all homologous recombination events were correct was confirmed by Southern blotting.

Pulse-chase experiments and immunoprecipitations. For pulse-chase experiments, 2.5 A_{600} cells were taken from a logarithmically growing culture for each time point and were labelled with 62.5 μCi [^{35}S]-methionine. Growth, labelling, chase conditions and other experimental procedures, such as cell lysis, immunoprecipitation and SDS-PAGE, were performed as described¹³.

Deglycosylation experiments and western analysis. Cells were grown at the indicated temperature in complete synthetic medium containing 2% glucose to an A_{600} of 3.0. Immunoprecipitation and deglycosylation of CPY* was performed as described⁵. Immunoprecipitated material was boiled in 50 μl UREA buffer before SDS-PAGE using a 8% gel and blotting. For western analysis, detection of the indicated proteins was performed using the respective antibodies.

Protease protection experiments. Spheroplasting and cell breakage were done as described⁵. For protease treatment of the pellet, trypsin was added to a final concentration of 0.5 mg ml^{-1} after resuspension of the pellet. The samples were incubated for 30 min on ice. If added, Triton X-100 was present at 1%. All treatments were stopped by TCA precipitation. After resuspending the pellet in 100 μl UREA buffer, CPY* was analysed by SDS-PAGE and immunoblotting.

β -Galactosidase activity test. After adding cycloheximide to a final concentration of 0.5 mg ml^{-1} at zero time ($t = 0$) to the logarithmically growing culture, 0.3 A_{600} of cells were taken for each time point, mixed with lysis buffer (0.6% Triton X-100, 0.75% ONPG, 2.25% β -ME, 0.15 M Tris-HCl, pH 7.5) and kept at -80°C for 30 min. After incubation for 60–90 min at 37°C , 75 μl of 1 M NaHCO_3 was added to the samples, debris was removed by centrifugation (20,000g, 3 min) and A_{405} was determined.

Received 14 July; accepted 31 July 1997.

1. Pryer, N. K., Wuestehube, L. J. & Schekman, R. Vesicle-mediated protein sorting. *Annu. Rev. Biochem.* **61**, 471–516 (1992).
2. Gething, M. J. & Sambrook, J. Protein folding in the cell. *Nature* **355**, 33–45 (1992).
3. Kopito, R. R. ER quality control: the cytoplasmic connection. *Cell* **88**, 427–430 (1997).
4. Biederer, T., Volkwein, C. & Sommer, T. Degradation of subunits of the Sec61p complex, an integral component of the ER membrane, by the ubiquitin-proteasome pathway. *EMBO J.* **15**, 2069–2076 (1996).
5. Hiller, M. M., Finger, A., Schweiger, M. & Wolf, D. H. ER degradation of a misfolded luminal protein by the cytosolic ubiquitin-proteasome pathway. *Science* **273**, 1725–1728 (1996).
6. Werner, E. D., Brodsky, J. L. & McCracken, A. A. Proteasome-dependent endoplasmic reticulum-associated protein degradation: an unconventional route to a familiar fate. *Proc. Natl Acad. Sci. USA* **93**, 13797–13801 (1996).
7. Jensen, T. J. *et al.* Multiple proteolytic systems, including the proteasome, contribute to CFTR processing. *Cell* **83**, 129–135 (1995).
8. Halaban, R. *et al.* Aberrant retention of tyrosinase in the endoplasmic reticulum mediates accelerated degradation of the enzyme and contributes to the dedifferentiated phenotype of amelanotic melanoma cells. *Proc. Natl Acad. Sci. USA* **94**, 6210–6215 (1997).
9. Walter, P. & Johnson, A. E. Signal sequence recognition and protein targeting to the endoplasmic reticulum membrane. *Annu. Rev. Cell Biol.* **10**, 87–119 (1994).
10. Rapoport, T. A., Jungnickel, B. & Kutay, U. Protein transport across the eukaryotic endoplasmic reticulum and bacterial inner membranes. *Annu. Rev. Biochem.* **65**, 271–303 (1996).
11. Wiertz, E. J. H. J. *et al.* Sec61-mediated transfer of a membrane protein from the endoplasmic reticulum to the proteasome for destruction. *Nature* **384**, 432–438 (1996).
12. Lyman, S. K. & Schekman, R. Binding of secretory precursor polypeptides to a translocon subcomplex is regulated by BiP. *Cell* **88**, 85–96 (1997).
13. Finger, A., Knop, M. & Wolf, D. H. Analysis of two mutated vacuolar proteins reveals a degradation pathway in the endoplasmic reticulum or a related compartment of yeast. *Eur. J. Biochem.* **218**, 565–574 (1993).
14. Stirling, C. S., Rothblatt, J., Hosobuchi, M., Deshaies, R. & Schekman, R. Protein translocation mutants defective in the insertion of integral membrane proteins into the endoplasmic reticulum. *Mol. Cell Biol.* **3**, 129–142 (1992).
15. Panzner, S., Dreier, L., Hartmann, E., Kostka, S. & Rapoport, T. A. Posttranslational protein transport in yeast reconstituted with a purified complex of Sec proteins and Kar2p. *Cell* **81**, 561–570 (1995).
16. Finke, K. *et al.* A second trimeric complex containing homologs of the Sec61p complex functions in protein transport across the ER membrane of *S. cerevisiae*. *EMBO J.* **15**, 1482–1494 (1996).
17. Feldheim, D., Rothblatt, J. & Schekman, R. Topology and functional domains of Sec63p, an endoplasmic reticulum membrane protein required for secretory protein translocation. *Mol. Cell Biol.* **12**, 3288–3296 (1992).
18. Corsi, A. K. & Schekman, R. The luminal domain of Sec63p stimulates the ATPase activity of BiP and mediates BiP recruitment to the translocon in *Saccharomyces cerevisiae*. *J. Cell Biol.* **137**, 1483–1493 (1997).
19. Cyr, D. M., Langer, T. & Douglas, M. G. DnaJ-like proteins: molecular chaperones and specific regulators of Hsp70. *Trends Biochem. Sci.* **19**, 176–181 (1994).
20. Hanein, D. *et al.* Oligomeric rings of the Sec61p complex induced by ligands required for protein translocation. *Cell* **87**, 721–732 (1996).
21. Scidmore, M. A., Okamura, H. H. & Rose, M. D. Genetic interactions between KAR2 and SEC63, encoding eukaryotic homologues of DnaK and DnaJ in the endoplasmic reticulum. *Mol. Biol. Cell* **4**, 1145–1159 (1993).
22. Sanders, S. L., Whitfield, K. M., Vogel, J. P., Rose, M. D. & Schekman, R. Sec61p and BiP directly facilitate polypeptide translocation into the ER. *Cell* **69**, 353–365 (1992).
23. Esnault, Y., Blondel, M.-O., Deshaies, R. J., Schekman, R. & Képès, F. The yeast SSS1 gene is essential for secretory protein translocation and encodes a conserved protein of the endoplasmic reticulum. *EMBO J.* **12**, 4083–4093 (1993).
24. Ng, D. T., Brown, J. D. & Walter, P. Signal sequences specify the targeting route to the endoplasmic reticulum membrane. *J. Cell Biol.* **134**, 269–278 (1996).

25. Chen, P., Johnson, P., Sommer, T., Jentsch, S. & Hochstrasser, M. Multiple ubiquitin-conjugating enzymes participate in the in vivo degradation of the yeast MAT $\alpha 2$ repressor. *Cell* **74**, 357–369 (1993).
26. Knop, M., Finger, A., Braun, T., Hellmuth, K. & Wolf, D. H. Der1, a novel protein specifically requires for endoplasmic reticulum degradation in yeast. *EMBO J.* **15**, 753–763 (1996).
27. Hampton, R. Y., Gardner, R. G. & Rine, J. Role of 26S proteasome and HRD genes in the degradation of 3-hydroxy-3-methylglutaryl-CoA reductase, an integral endoplasmic reticulum membrane protein. *Mol. Biol. Cell* **7**, 2029–2044 (1994).
28. Ausubel, F. M. *et al.* *Current Protocols in Molecular Biology* (Greene, New York, 1992).
29. Rose, M. D., Misra, L. M. & Vogel, J. P. Kar2, a karyogamy gene, is the yeast homolog of the mammalian BiP/GRP78 gene. *Cell* **57**, 1211–1221 (1989).

Acknowledgements. We thank M. Hochstrasser (Deg1- β -galactosidase fusion), M. D. Rose (kar2-113 allele), T. A. Rapoport (Sec61p antibodies), H. K. Rudolph and R. Schekman (Kar2 antibodies), M. Knop (polyclonal CPY antibodies) and S. Rupp (proteinase yscA antibodies) for providing gene constructs and affinity purified antibodies; and S. Jäger and M. Hammerle for discussions. This work was supported by the Bundesministerium für Forschung und Technologie and the Fonds der Chemischen Industrie, Frankfurt.

Correspondence and requests for materials should be addressed to D.H.W. (e-mail: dieter.wolf@po.uni-stuttgart.de).

Marking of active genes on mitotic chromosomes

Emil F. Michelotti*, Suzanne Sanford & David Levens

Gene Regulation Section, Laboratory of Pathology, National Cancer Institute, National Institutes of Health, Building 10, Room 2N105, Bethesda, Maryland 20892, USA

During development and differentiation, cellular phenotypes are stably propagated through numerous cell divisions¹. This epigenetic 'cell memory' helps to maintain stable patterns of gene expression². DNA methylation³ and the propagation of specific chromatin structures may both contribute to cell memory⁴. There are two impediments during the cell cycle that can hinder the inheritance of specific chromatin configurations: first, the pertinent structures must endure the passage of DNA-replication forks in S phase⁵; second, the chromatin state must survive mitosis, when chromatin condenses, transcription is turned off, and almost all double-stranded DNA-binding proteins are displaced^{6,7}. After mitosis, the previous pattern of expressed and silent genes must be restored. This restoration might be governed by mass action, determined by the binding affinities and concentrations of individual components. Alternatively, a subset of factors might remain bound to mitotic chromosomes, providing a molecular bookmark to direct proper chromatin reassembly. Here we analyse DNA at transcription start sites during mitosis *in vivo* and find that it is conformationally distorted in genes scheduled for reactivation but is undistorted in repressed genes. These protein-dependent conformational perturbations could help to re-establish transcription after mitosis by 'marking' genes for re-expression.

DNase I-hypersensitive sites in chromatin are useful indicators of gene activity and may persist during mitosis, despite cessation of transcription and dissociation of most transcription factors from mitotic chromosomes⁶. Many DNase I-hypersensitive sites are also sensitive to S1 nuclease and hence may have single-stranded features^{8–10}. Melted regions of the human *c-myc* gene have been mapped in unsynchronized cells and shown to bind transcription factors *in vitro*^{10,11}. During mitosis, the single-stranded properties of chromatin increase ~ 10 -fold¹². To test whether any mitosis-specific increase in single-stranded character might occur in previously characterized S1-sensitive regions of the human *c-myc* gene, bases reactive to potassium permanganate^{10,11} were mapped by using ligation-mediated polymerase chain reaction (LM-PCR). Factors binding to such melted regions during mitosis may contribute to the inheritance of the *c-myc* gene chromatin structure through this

* Present address: Genelabs Technologies, 505 Penobscot Drive, Redwood City, California 94043, USA

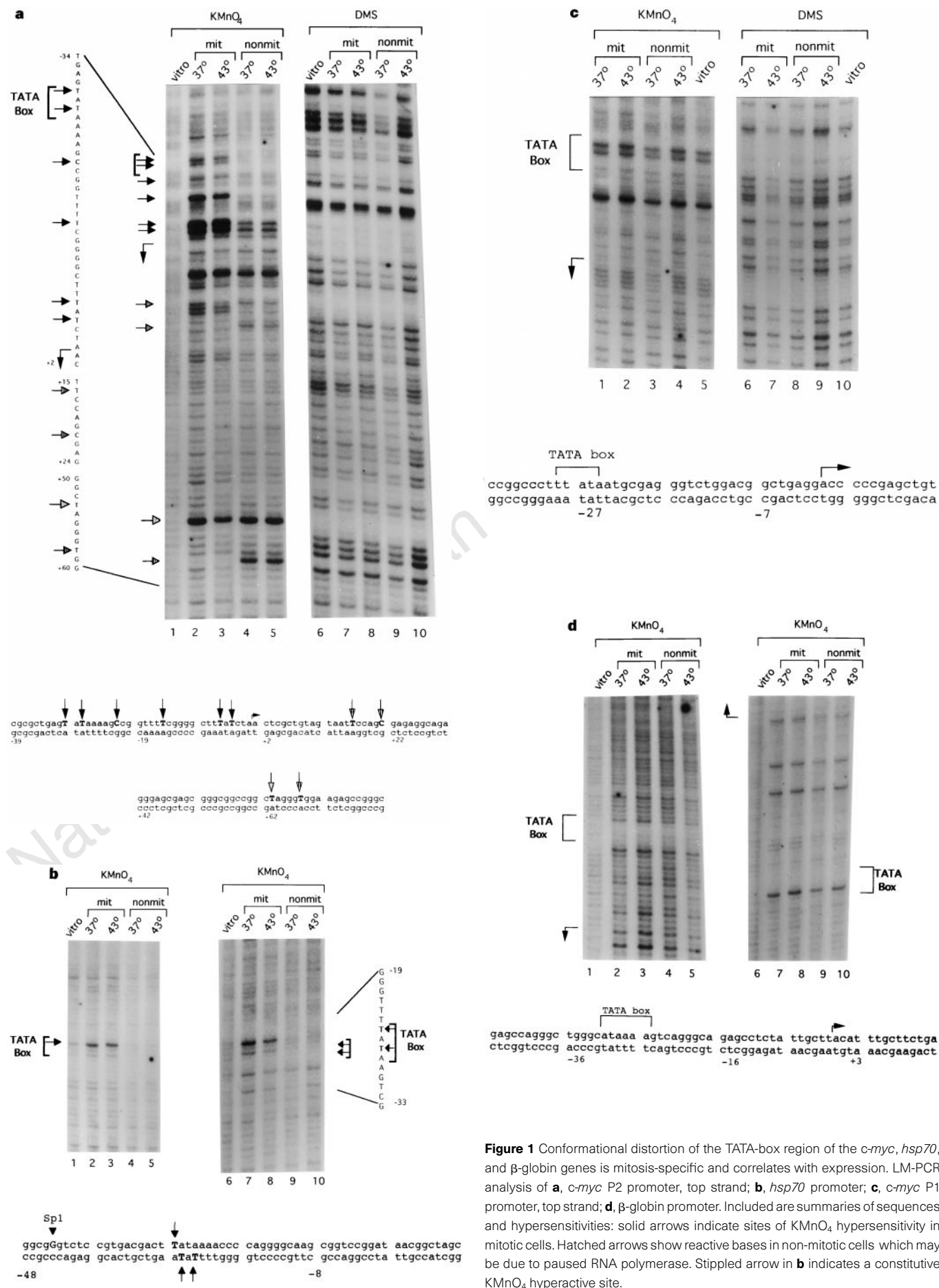


Figure 1 Conformational distortion of the TATA-box region of the *c-myc*, *hsp70*, and β -globin genes is mitosis-specific and correlates with expression. LM-PCR analysis of **a**, *c-myc* P2 promoter, top strand; **b**, *hsp70* promoter; **c**, *c-myc* P1 promoter, top strand; **d**, β -globin promoter. Included are summaries of sequences and hypersensitivities: solid arrows indicate sites of KMnO₄ hypersensitivity in mitotic cells. Hatched arrows show reactive bases in non-mitotic cells which may be due to paused RNA polymerase. Stippled arrow in **b** indicates a constitutive KMnO₄ hyperactive site.

phase of the cell cycle. This approach has identified and mapped melted DNA at promoters bearing paused RNA polymerases¹³ and at other sites binding single-strand-specific factors^{10,11}. The *c-myc* P2 promoter, a major region of S1 sensitivity in isolated nuclei, was analysed with KMnO_4 during mitosis in synchronized cells. In parallel, *in vivo* dimethylsulphate (DMS) footprinting of the *hsp70* promoter was used to monitor displacement of conventional double-stranded DNA binding proteins⁶. HeLa cells were treated with nocodazole for eight hours⁶ (which synchronizes cells at metaphase), separated by shake-off into mitotic and non-mitotic cells, incubated for 30 min at either 37 °C or 43 °C (for analysis of the *hsp70* promoter), and treated *in vivo* with either DMS or KMnO_4 (ref. 9). Purified DNA was subjected to LM-PCR using primers that allowed visualization of the top strand of the human *c-myc* gene in the vicinity of the P2 transcription start site.

KMnO_4 reactions revealed unexpected mitosis-specific unwinding of the *myc* P2 TATA-box/start-site region (Fig. 1a; compare lanes 2, 3 with 4, 5). These changes were not apparent by DMS footprinting (Fig. 1a; compare lanes 7, 8 with 9, 10). Similar changes were noted in mitotic cells prepared without nocodazole and hence were not caused by blocking the cell cycle (data not shown). Controls confirmed that cells were properly synchronized (mitotic index of >95%), that dramatic mitotic displacement of the conventional double-stranded DNA-binding proteins Sp1 and HSF (heat-shock factor) from the *hsp70* promoter proceeded as described⁶, and that upregulation of *in vivo* binding by HSF (and Sp1 to a lesser extent) occurred upon heat shock only in non-mitotic cells (data not shown).

The *hsp70* TATA-box region⁶ was examined during mitosis. Like the *c-myc* P2 promoter, the *hsp70* TATA box experienced a conformational change that was specific for mitosis (Fig. 1b; compare lanes 2, 3 with 4, 5; and lanes 7, 8 with 9, 10). Again, these changes were not seen with DMS (data not shown). Not all TATA boxes

were mitotically distorted, because no changes occurred at *c-myc* P1 (Fig. 1c; compare lanes 1, 2 with 3, 4), which also appeared to be vacant as judged by DMS footprinting. Perhaps the conformational changes at the *hsp70* and *c-myc* promoters marked these genes for transcriptional restart following mitosis. If so, the proximal promoter region of non-expressed genes should be unreactive with KMnO_4 during mitosis. As predicted, neither strand of the silent β -globin gene showed mitosis-specific changes (Fig. 1d, compare lanes 2, 3 with 4, 5, and lanes 7, 8 with 9, 10). Thus, stable gene activity correlated with perturbation of the TATA box during mitosis.

If mitotic TATA-box marking were linked to the maintenance of gene expression, then no unwinding at the *c-myc* P2 TATA-box region should occur when *c-myc* transcription is repressed. In IMR32 neuroblastoma cells overexpressing N-myc, the *c-myc* locus is shut off and *c-myc* RNA is undetectable¹⁰. No change in DNA conformation of the *c-myc* P2 proximal promoter region was detected with KMnO_4 (Fig. 2a) or DMS (Fig. 2a, right), either in mitotic or non-mitotic IMR32 cells (Fig. 2a; compare lanes 6, 7 with 2, 3), indicating that mitosis-specific tagging of this start site was linked to gene expression. Likewise, no KMnO_4 hypersensitivity was induced on the complementary strand (Fig. 2a, right). DMS reactions in HeLa and IMR32 cells were indistinguishable (Fig. 2a, lanes 8–12). The lymphoma cell line CA46, which expresses *c-myc*, was also mitotically marked at P2 (Fig. 2a, lanes 4 and 5).

The lack of mitosis-specific marking of *c-myc* in IMR32 cells might simply have been due to a cell-line-specific defect in whatever machinery was required for unwinding. We therefore tested the mitotic KMnO_4 reactivity of the *hsp70* gene, which is expressed in both IMR32 and HeLa cells. The *hsp70* TATA-box region was similarly tagged in neuroblastoma cells (Fig. 2b, lanes 1 and 2) and HeLa cells (Fig. 2b, lanes 3 and 4), indicating that the mechanism for TATA-box unwinding is intact in IMR32 cells; therefore the

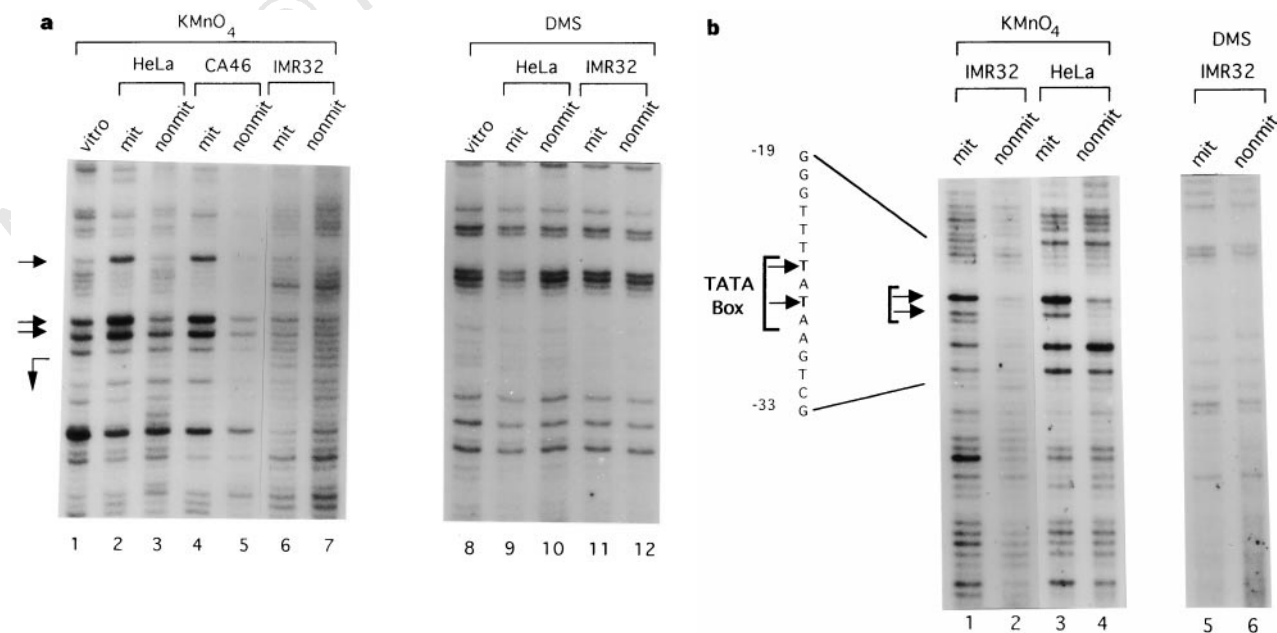


Figure 2 a, The *c-myc* P2 TATA-box region is not conformationally perturbed during mitosis when not expressed in IMR32 neuroblastoma cells. Genomic DNA *in vitro* or intact synchronized HeLa, CA46, or IMR32 cells were treated with KMnO_4 (lanes 1–7) or DMS (lanes 8–12) and analysed by LM-PCR. **b**, the *hsp70* TATA box region is mitotically marked in IMR32 neuroblastoma cells. DNA reacted *in vivo* with KMnO_4 (lanes 1–4) or DMS (lanes 5, 6) from synchronized IMR32 neuroblastoma and HeLa cells was used to assess marking of the bottom strand of the *hsp70* promoter region via LM-PCR.

lack of unwinding at the *c-myc* P2 promoter in IMR32 was probably linked to the absence of *c-myc* transcription.

Genes lacking TATA boxes, including some 'housekeeping' genes, must also be restarted after mitosis. If mitotic marking contributes generally to re-expression, then expressed TATA-less genes such as the *ets-2* protooncogene¹⁴ should display mitotic perturbation near their initiation sites. Indeed, several nucleotides at the *ets-2* start site and in the adjacent upstream sequence were mitotically reactive with KMnO_4 (Fig. 3, arrows), whereas downstream residues were reactive only in non-mitotic cells (Fig. 3, arrowheads), which is consistent with transcriptional melting. Furthermore, a TATA-less promoter that is silent in HeLa cells, that for the terminal deoxynucleotidyl transferase (TdT) gene^{15,16} was devoid of mitosis-specific characteristics (data not shown).

What are the molecular requirements for placement of mitotic bookmarks? The diversity of affected promoters and the association with gene activity argues strongly against specific sequence recognition as the basis for mitotic marking. The correlation of mitotic tags with the activity of the *c-myc*, *hsp70*, *ets-2*, TdT or β -globin genes indicates that the conformational distortion is not due solely to chromosome condensation. However, the mitotic mark might represent buckling of naked promoter DNA, vacated during metaphase, under mechanical stress, but such a mechanism should have strained both *c-myc* promoters equally. Alternatively, melting might require a protein factor or cofactor. If this hypothetical protein's half-life were sufficiently short, then the metaphase-specific KMnO_4 hyperactive sites at the *c-myc* P2 promoter should decay with cycloheximide treatment. Therefore, nocodazole-synchronized mitotic HeLa cells were divided into two aliquots and incubated for varying lengths of time, with or without $25 \mu\text{g ml}^{-1}$ cycloheximide. Cycloheximide- and mock-treated cells were reacted *in vivo* in the presence of nocodazole with KMnO_4 for LM-PCR analysis. A cycloheximide-induced decrease in mitosis-specific KMnO_4 reactivity (relative to mock-treated cells) at the *c-myc* P2 promoter was first detected 3 h after cycloheximide treatment (Fig. 4a; compare lanes 5 and 6 with 3 and 4) and was clearly evident at 4 h (Fig. 4a, lanes 7 and 8). Also potentially turning over in this experiment, however, were proteins required for chromosome condensation; the resulting chromosome decompaction could compromise interpretation of *in vivo* KMnO_4 analysis. To exclude these complications, both mock- and cycloheximide-treated cells from the 4-h time point were fixed, stained¹⁷ and microscopically examined; no changes in chromosome morphology induced by cycloheximide treatment were evident (data not shown). This experiment supports the existence of a protein required for the maintenance of the mitosis-specific mark at the *c-myc* P2 start site.

These results provide *in vivo* evidence that a protein is required for mitosis-specific melting of the *c-myc* P2 start site. *In vitro* evidence for such a protein was sought by testing the salt-extractability of this factor from purified mitotic chromosomes. Nocodazole-arrested mitotic HeLa cells were lysed in hypotonic buffer, condensed chromosomes were centrifugally purified and then extensively washed with buffers of increasing ionic strength. Washed mitotic chromosomes were then treated with KMnO_4 and analysed by LM-PCR. The mitotic mark resisted extraction with 10 mM NaCl (Fig. 4b, lane 1), 200 mM NaCl (Fig. 4b, lane 2) and 300 mM NaCl (data not shown). Extraction with 400 mM NaCl yielded a pattern of KMnO_4 hypersensitive site very similar to that of non-mitotic cells (Fig. 4b; compare lanes 3 and 4). A mitotic marker was extracted, but 400 mM NaCl is also sufficient to mobilize nucleosomes and potentially alter other features peculiar to the mitotic chromatin of active genes.

Re-addition of the 0.4 M NaCl extract to the salt-washed mitotic chromosomes failed to restore the mitotic marker. It is unlikely that simple sequence recognition mediates direct formation of marker complexes at active promoters. If the marker itself instructs genes to

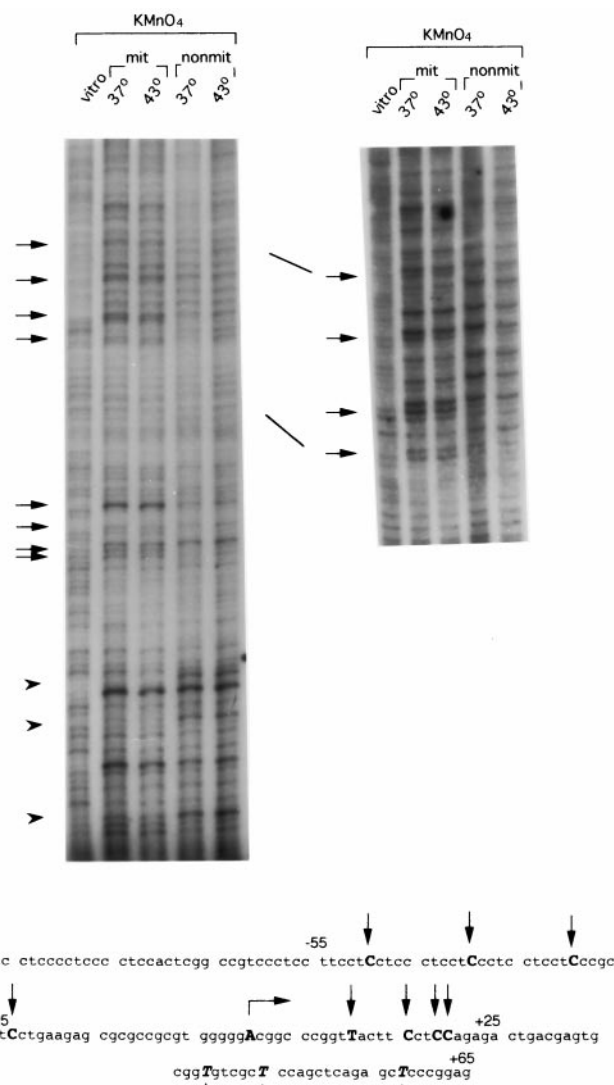


Figure 3 The *ets-2* start site is marked during mitosis. Left, bases mitotically reactive with KMnO_4 are indicated with arrows; arrowheads indicate bases selectively reactive in non-mitotic cells. Right, longer run of sequences upstream of -20. Bottom, summary of reactive bases; bent arrow indicates the start-site (+1).

be post-mitotically expressed, then removal of the marker should convert expressed genes to the default silent state. In the absence of transcriptional activity, reconstitution of the marker would not occur. Alternatively, the mitotic marker may act as an accessory factor to maintain a segment of DNA already configured for transcription, ready for reoccupancy by the transcription machinery. In this scheme, accurate reconstitution of the mitotic marker would be achievable unless the salt treatment disrupted native mitotic chromatin or removed cofactors.

The mitosis-specific KMnO_4 hypersensitivity indicates a nidus of single-stranded DNA at start sites, perhaps stabilized by bound protein. Two features of single-stranded DNA might facilitate transcription restart. First, nucleosomes, which might otherwise obstruct the promoter, have a reduced affinity for single-stranded DNA¹⁸. Second, the deformability of single-stranded DNA to flexural and torsional stress¹⁹ facilitates the assembly of the transcription complex, especially if the returning TFIID or other basal components recognized the bookmarker⁷. *In vivo* footprinting indicates that other single-stranded DNA-binding proteins, hnRNP

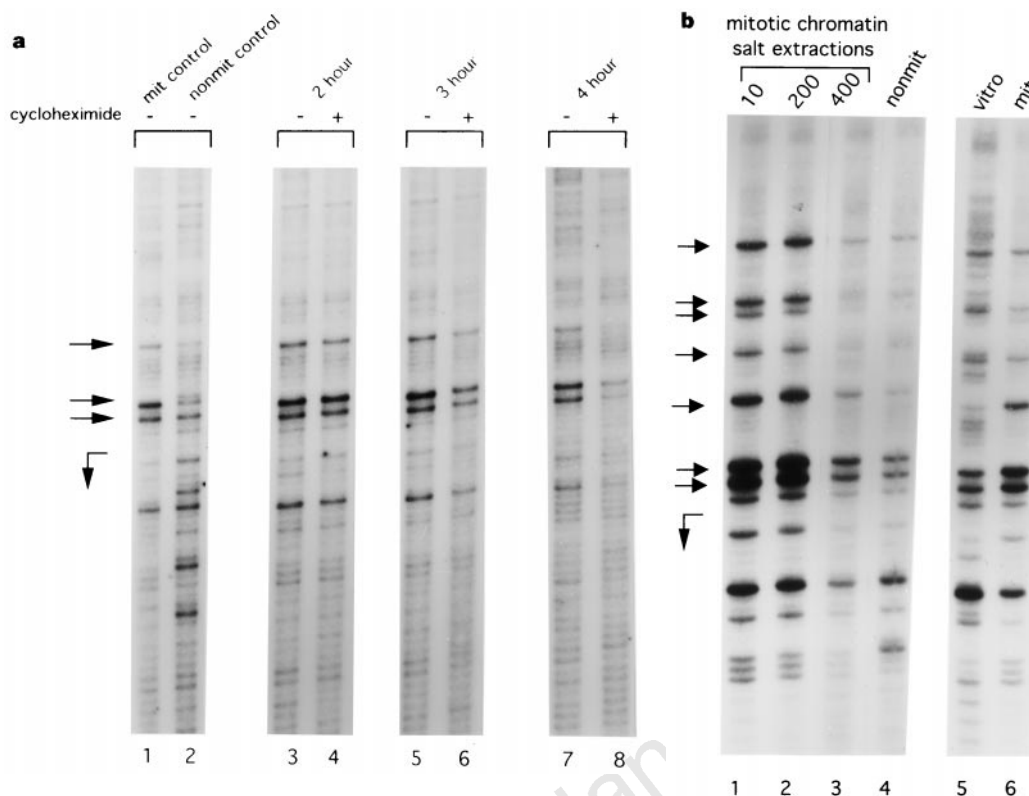


Figure 4 Protein requirement for the melting of the *c-myc* P2 TATA-box region during mitosis. **a**, Nocodazole-arrested mitotic HeLa cells were shaken off and separated into two nocodazole-containing flasks, one with cycloheximide, one without cycloheximide. At various times, samples were treated *in vivo* with KMnO_4 , and purified DNA was analysed by LM-PCR using primers that visualized the top strand of the *c-myc* P2 promoter. **b**, Factor(s) required for mitosis-specific *c-myc* P2 promoter KMnO_4 hypersensitivity is salt-extractable. HeLa cells were treated with nocodazole for 8 h and mitotic cells were shaken off. Cells were then hypotonically lysed and mitotic chromosomes purified by centrifugation. Chromosomes were then extracted with increasing concentrations of NaCl, and treated with KMnO_4 . The top strand of the *c-myc* P2 promoter was then analysed by LM-PCR.

K and FBP⁹, remain bound to the human *c-myc* promoter during mitosis (data not shown) when the single-stranded character of bulk chromatin increases 10-fold¹². Single-stranded DNA and associated factors may therefore be important for mitosis.

What is the source of single-stranded DNA during mitosis? Possibly transcription; melted DNA at promoter start sites in this scheme is stabilized by a factor, perhaps also interacting with the transcription machinery, just before M phase. Persistence of the marker through mitosis may require sustained torque. The requirement for topoisomerase I activity in chromosome condensation indicates that torsional strain is indeed present during compaction²⁰. Recruitment of the cellular transcription machinery to genes marked by stressed DNA could be important to re-establish the correct range of expressed genes. □

Methods

Ligation-mediated PCR. *In vivo* footprinting was done as described²¹. Primers for LM-PCR analysis of the top strand of the *c-myc* P2 region were: TAGC-CCCCTATTCGCTC, CCCCTATTCGCTCCGATCTCCC and TTGCTCCGGATCTCCCTTCCCAGGA. Primers for LM-PCR analysis of the top strand of *c-myc* P1 were: TCTCGAGGCAGGAGGGG, CGAGGCAGGAGGGGAGC-CAGGGAC and GGGAGGAGGGGAGCCAGGGACGGCCGG. Primers for analysis of the bottom strand of the globin TATA-box region were: CACT-TAGACCTCACCCTG, GACCTCACCCTGTCCAGCCACAC and CTGT-GGAGCCACACCCTAGGGTTGGCC; primers for the top strand of globin were: GTCAGGTGCACCATGGTG, TGCACCATGGTGTCTGTTTGGAGG and GTGTCTGTTTGGAGTTGCTAGTGAAC. Primers for the strand of the *ets-2* promoter were: CTCCGGGAGCTCTGAGC, CGGGAGCTCTGAGCTG-GAGCGAC and GAGCTCTGAGCTGGAGCGACACCGCA.

LM-PCR analysis of mitotic chromosomes. Washed HeLa cells were Dounce-homogenized in 10 mM Tris, pH 8.0, 10 mM KCl. Mitotic chromosomes were pelleted by a 2-min spin in the microfuge and washed with a buffer containing 10 mM Tris, pH 8.0, 1.5 mM MgCl_2 , 0.5 mM EDTA, plus the indicated amounts of salt. LM-PCR analysis then proceeded as with intact cells. Fixing and staining of mitotic chromosomes was done as before¹⁷.

Received 27 January; accepted 9 June 1997.

- Alberts, B. *et al.* *Molecular Biology of the Cell* (Garland, New York and London, 1994).
- Weintraub, H. Assembly and propagation of repressed and depressed chromosomal states. *Cell* **42**, 705–711 (1985).
- Martienssen, R. A. & Richards, E. J. DNA methylation in eukaryotes. *Curr. Opin. Genet. Dev.* **5**, 234–242 (1995).
- Groudine, M. & Weintraub, H. Propagation of globin DNAase I-hypersensitive sites in absence of factors required for induction: a possible mechanism for determination. *Cell* **30**, 131–139 (1982).
- Liu, B. & Alberts, B. M. Head-on collision between a DNA replication apparatus and RNA polymerase transcription complex. *Science* **267**, 1131–1137 (1995).
- Martinez-Balbas, M. A., Dey, A., Rabindran, S. K., Ozato, K. & Wu, C. Displacement of sequence-specific transcription factors from mitotic chromatin. *Cell* **83**, 29–38 (1995).
- Segil, N., Guermah, M., Hoffmann, A., Roeder, R. G. & Heintz, N. Mitotic regulation of TFIID: inhibition of activator-dependent transcription and changes in subcellular localization. *Genes Dev.* **10**, 2389–2400 (1996).
- Larsen, A. & Weintraub, H. An altered DNA conformation detected by S1 nuclease occurs at specific regions in active chick globin chromatin. *Cell* **29**, 609–622 (1982).
- Michelotti, E. F., Michelotti, G. A., Aronson, A. & Levens, D. Heterogeneous nuclear ribonucleoprotein K is a transcription factor. *Mol. Cell. Biol.* **16**, 2350–2360 (1996).
- Michelotti, G. A. *et al.* Multiple single-stranded *cis* elements are associated with activated chromatin of the human *c-myc* gene *in vivo*. *Mol. Cell. Biol.* **16**, 2656–2669 (1996).
- Duncan, R. *et al.* A sequence-specific, single-strand binding protein activates the far upstream element of *c-myc* and defines a new DNA binding motif. *Genes Dev.* **8**, 465–480 (1994).
- Juan, G., Pan, W. & Darzynkiewicz, Z. DNA segments sensitive to single-strand-specific nucleases are present in chromatin of mitotic cells. *Exp. Cell Res.* **227**, 197–202 (1996).
- Giardina, C., Perez-Riba, M. & Lis, J. T. Promoter melting and TFIID complexes on *Drosophila* genes *in vivo*. *Genes Dev.* **6**, 2190–2200 (1992).
- Mavrothalassitis, G. W., Watson, D. K. & Papas, T. The human *ETS-2* gene promoter: molecular dissection and nuclease hypersensitivity. *Oncogene* **5**, 1337–1342 (1990).
- Koiwai, O. & Morita, A. Isolation of a putative promoter region for human terminal deoxynucleotidyltransferase gene. *Biochem. Biophys. Res. Commun.* **154**, 91–100 (1988).
- Bhaumik, D. *et al.* Identification of a tripartite basal promoter which regulates human terminal deoxynucleotidyl transferase gene expression. *J. Biol. Chem.* **269**, 15861–15867 (1994).
- Schweizer, D. & Ambros, P. F. *Chromosome Analysis Protocols* 1-97-112 (Humana, Totowa, 1994).
- Bonne-Andrea, C., Wong, M. L. & Alberts, B. M. *In vitro* replication through nucleosomes without histone displacement. *Nature* **343**, 719–726 (1990).
- Kahn, J. D., Yun, E. & Crother, D. M. Detection of localized DNA flexibility. *Nature* **368**, 163–166 (1994).
- Castano, I. B., Brzoska, P. M., Sadoff, B. U., Chen, H. & Christman, M. F. Mitotic chromosome condensation in the rDNA requires TRF4 and DNA topoisomerase I in *Saccharomyces cerevisiae*. *Genes Dev.* **10**, 2546–2576 (1996).
- Mueller, P. R. & Wold, B. *In vivo* footprinting of a muscle specific enhancer by ligation mediated PCR. *Science* **246**, 780–786 (1989).

Acknowledgements. We thank L. Liotta, C. Wu, S. Mackem and M. Reitman for critical review of the manuscript.

Correspondence and requests for materials should be addressed to D.L.

Silencing factors participate in DNA repair and recombination in *Saccharomyces cerevisiae*

Yasumasa Tsukamoto, Jun-ichi Kato & Hideo Ikeda

Department of Molecular Biology, Institute of Medical Science, University of Tokyo, P.O. Takanawa, Tokyo 108, Japan

DNA double-strand breaks are repaired by homologous recombination or DNA end-joining, but the latter process often causes illegitimate recombination and chromosome rearrangements. One of the factors involved in the end-joining process is Hdf1, a yeast homologue of Ku protein¹⁻⁴. We used the yeast two-hybrid assay to show that Hdf1 interacts with Sir4, which is involved in transcriptional silencing at telomeres and *HM* loci^{5,6}. Analyses of *sir4* mutants showed that Sir4 is required for deletion by illegitimate recombination and DNA end-joining in the pathway involving Hdf1. Sir2 and Sir3, but not Sir1, were also found to participate in these processes. Furthermore, mutations of the *SIR2*, *SIR3* and *SIR4* genes conferred increased sensitivity to γ -radiation in a genetic background with a mutation of the *RAD52* gene, which is essential for double-strand break repair mediated by homologous recombination. These results indicate that Sir proteins are involved in double-strand break repair mediated by end-joining. We propose that Sir proteins act with Hdf1 to alter broken DNA ends to create an inactivated chromatin structure that is essential for the rejoining of DNA ends.

Illegitimate recombination occurs between DNA sequences that are non-homologous or have very short regions of homology, and is the result of end-joining during the repair of DNA double-strand breaks. We have studied the formation of deletions resulting from illegitimate recombination in the yeast *Saccharomyces cerevisiae*, and found that the proteins Rad50, Mre11, Xrs2 and Hdf1 are involved in illegitimate recombination and end-joining^{1,2,7}. Hdf1 is a yeast homologue of the 70K subunit of the mammalian Ku protein, which can bind to the ends of double-stranded DNA and mediates double-strand break repair and immunoglobulin V(D)J recombination^{8,9}. The circularization of linear DNA by end-joining is less efficient in the yeast *hdf1* mutant and in cells with a mutation in the *KU80* gene, which codes for a homologue of another subunit of the Ku protein^{3,4,10}. The *hdf1 rad52* and *ku80 rad52* double mutants are more sensitive to γ -radiation than the isogenic *rad52* single mutant, suggesting that most DNA double-strand breaks are repaired by end-joining in the absence of repair mediated by homologous recombination^{4,10,11}. Mammalian *xrs6* mutant cells, which have a defect in one of the subunits of Ku protein, exhibit hypersensitivity to X-rays and are deficient in end-joining but not homologous recombination, indicating that end-joining is important for DNA double-strand break repair in mammalian cells¹².

To investigate the function of Hdf1 in end-joining, we used the yeast two-hybrid system¹³ to search for yeast proteins that interact with Hdf1. Plasmid pLexHDF1, which carries the *HDF1* gene fused to the DNA-binding-domain of LexA, was constructed and used for screening. From a library of genes fused to the Gal4 activation domain, 42 positive clones were isolated. Sequence analysis revealed that 11 of these clones had plasmids with the *GAL4* fused in-frame to the *SIR4* gene, which codes for a silencing information regulator. The Sir4 part of all fusions was found to begin at amino acid 1,205 of Sir4, and polymerase chain reaction (PCR) analysis confirmed that the sequence continued to the carboxy-terminal amino acid 1,358. The *GAL4-SIR4* plasmid induced specific expression of the *lacZ* reporter gene in the presence of the *lexA-HDF1* plasmid (Table 1). Sir4 is a regulatory factor in transcriptional silencing at telomeres

and *HM* mating-type loci in yeast^{5,6}. In addition to Sir4, the proteins Sir2, Sir3, Rap1 and histones H3 and H4 are necessary for silencing at telomeres^{5,6}, and are thought to function by forming an inactivated chromatin structure analogous to the heterochromatin of higher eukaryotes^{5,6}.

To study the role of Sir4 in illegitimate recombination, we first examined the effect of the *sir4* mutation on deletion using a YCp plasmid, YCpL2 (ref. 7). YCpL2 carries two negative selection markers, the *CAN1* and *CYH2* genes, and three positive selection markers, the *TRP1*, *LEU2* and *URA3* genes. A *can1 cyh2* strain containing YCpL2 exhibits sensitivity to both canavanine and cycloheximide, and deletion on the *CAN1* and *CYH2* genes on YCpL2 can be detected by selecting Ura⁺ transformants resistant to both drugs. By structural analysis of the deleted plasmids, we have previously shown that deletion is caused by illegitimate recombination⁷. YCpL2 was introduced into a *sir4* mutant and the rate of illegitimate recombination was determined. The rate was reduced 20-fold in the *sir4* mutant compared to that of the isogenic *SIR4* strain (Fig. 1a). We also tested the effects on deletion of

Table 1 Interaction of Hdf1 and Sir4

DNA-binding domain plasmid	Activation domain plasmid	β -Gal activity (s.d.) (nmol min ⁻¹ mg protein ⁻¹)
pBTM116 (LexA)	pGAD424 (Gal4 _{AD})	1.1 (0.3)
pBTM116 (LexA)	pGADSIR4 (Gal4 _{AD} -Sir4)	1.7 (0.4)
pLexHDF1 (LexA-Hdf1)	pGAD424 (Gal4 _{AD})	0.2 (0.1)
pLexHDF1 (LexA-Hdf1)	pGADSIR4 (Gal4 _{AD} -Sir4)	20.9 (3.8)

The protein-protein interaction by two-hybrid assay was monitored by measuring the β -galactosidase activity of each L40 transformants²⁸. Results are an average of three determinations.

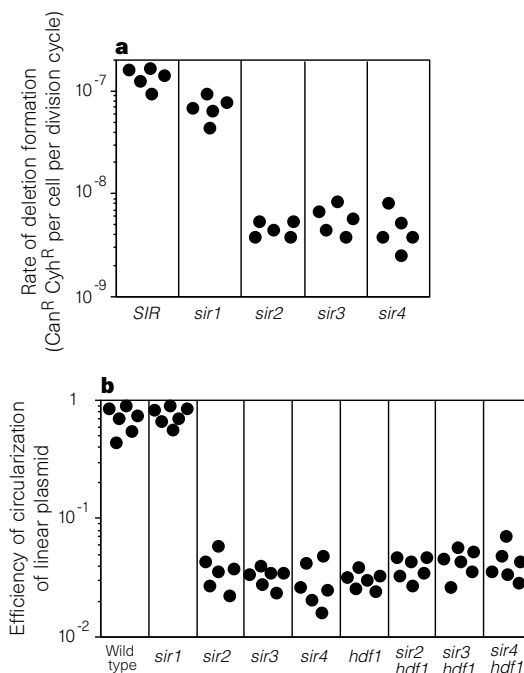


Figure 1 Deficiency in deletion formation and end-joining in *sir* mutants. **a**, Rate of deletion formation on the plasmid YCpL2. The recombination rate was determined by fluctuation analysis and data were analysed by the method of the median^{29,30}. **b**, The efficiency of end-joining of a linearized plasmid DNA. End-joining efficiency was calculated on the basis of the ratio of pRS316 (*Bam*HI digested; *URA3*) to YEpl3 (uncut; *LEU2*) transformants.

mutations in the *SIR1*, *SIR2* and *SIR3* genes. The rate of deletion formation was reduced in the *sir2* and *sir3* mutants, but not in the *sir1* mutant, indicating that Sir2, Sir3 and Sir4, but not Sir1, are involved in illegitimate recombination.

To determine the function of Sir proteins in illegitimate recombination, we then used a dicentric plasmid to examine the effects of *sir* mutations on deletion. A dicentric plasmid is structurally unstable in yeast, and DNA double-strand breaks are thought to occur frequently on this plasmid as a result of the two resident centromeres being pulled to opposite poles during mitosis. It has been shown that deletion occurs at a high frequency, and requires functions implicated in end-joining². Therefore, the frequency of deletions on the dicentric plasmid can be used as an indicator of the efficiency of end-joining. Because the dicentric plasmid YdCp2 has the *CAN1* and *CYH2* genes, deletions can be detected by the loss of these negative selection markers². The *sir1*, *sir2*, *sir3* and *sir4* mutants were transformed by YdCp2 and the frequency of deletions on the YdCp2 plasmid was measured (Table 2). The frequency was reduced in the *sir2*, *sir3* and *sir4* mutants, as well as in the *hdf1* mutant, both not in the *sir1* mutant, indicating that Sir2, Sir3 and Sir4, but not Sir1, are involved in end-joining. Moreover, the frequency of deletions in the *sir4 hdf1* double mutant was similar to that seen in the *sir4* or *hdf1* single mutant, suggesting that Sir4 and Hdf1 function in the same pathway during the joining of broken DNA ends.

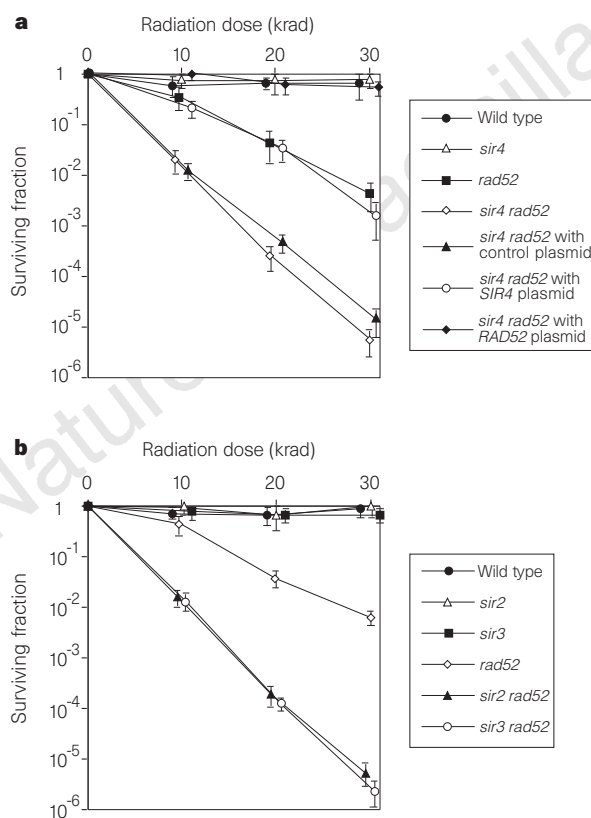


Figure 2 Sensitivity of *sir* and *rad52* double mutants to γ -irradiation. **a**, Survival of the *sir4* and *rad52* single mutants and the *sir4 rad52* double mutant after irradiation with γ -radiation. The control plasmid was the yeast vector pRS315 (*ARSH4 CEN6 LEU2*; ref. 14), and the *SIR4* and *RAD52* plasmids were the pRS315 containing the wild-type *SIR4* and *RAD52* genes, respectively. **b**, Survival of the *sir2*, *sir3* and *rad52* single mutants and the *sir2 rad52* and *sir3 rad52* double mutants after γ -irradiation. W303-1A and isogenic derivatives were used in this experiment. Error bars represent one standard error. Results are an average of five independent determinations.

To confirm the involvement of Sir protein function in end-joining, we examined the efficiency of circularization of linear DNA fragments introduced into these *sir* mutants. *Bam*HI digests of a *URA3* plasmid, pRS316 (ref. 14), and an uncut *LEU2* plasmid, YEp13 (ref. 15), were introduced into the *ura3 leu2* strain. Because there is no homology, the linearized pRS316 plasmid cannot be recircularized by homologous recombination with chromosomal DNA, or with the uncut YEp13 plasmid. The efficiency of circularization was estimated on the basis of the ratio of the number of *Ura*⁺ colonies to that of *Leu*⁺ colonies. In *sir2*, *sir3*, *sir4* and *hdf1* mutants, the efficiency was reduced 20-fold compared with the wild-type strain, whereas the efficiency was about the same in the *sir1* mutant (Fig. 1b). The deficiency of recircularization in the *sir2*, *sir3* and *sir4* mutants was complemented by the wild-type *SIR2*, *SIR3* and *SIR4* genes, respectively (data not shown). These results indicate that Sir2, Sir3 and Sir4 are essential for the rejoining of broken DNA ends, as shown previously for Hdf1. Because this result was consistent with the results of illegitimate recombination described above, the functions of Sir proteins involved in the end-joining step would be same as those for illegitimate recombination. The efficiency of circularization in the *sir2 hdf1*, *sir3 hdf1* and *sir4 hdf1* double mutants was comparable to that in the *sir2*, *sir3*, *sir4* or *hdf1* single mutants (Fig. 1b), confirming that Sir2, Sir3 and Sir4 work in the same pathway as Hdf1.

To investigate the effect of the *sir4* mutation on the structure of end-joining products, the recombinant plasmids rescued from *Ura*⁺ colonies of the wild-type or the *sir4* mutant were examined by restriction enzyme analysis. Of the 18 recombinant plasmids from the wild-type strain, 17 retained the *Bam*HI site, whereas only 3 of the 17 recombinant plasmids derived from the *sir4* mutant did so, indicating that most of the products from the *sir4* mutant contain deletion mutations around the *Bam*HI site. Analysis of the nucleotide sequence indicated that the recombinant plasmids derived from the *sir4* mutant had deletions at various sites, and there were only short regions of homology (2–5 base pairs) at the recombination junctions (data not shown). These results indicate that the *sir4* mutation reduces the frequency of precise end-joining in accordance with those with the *hdf1* and *ku80* mutations^{3,4,10}. Consistent with a previous study⁴, the efficiency of circularization of the linearized pRS316 plasmids with blunt ends was about 10-fold lower than those with staggered ends in the wild-type strain (data not shown), indicating that DNA ends with short complementary sequences are effectively joined to each other in the Sir4-dependent end-joining pathway.

To investigate whether Sir proteins are involved in the repair of DNA double-strand breaks, we examined the sensitivity of the *sir* mutants to γ -radiation. Although the *sir4* single mutant did not show increased sensitivity to ionizing radiation, the *sir4 rad52* double mutant was more sensitive than the *rad52* single mutant, and the radiation sensitivity of the *sir4 rad52* double mutant was complemented by the plasmids containing the wild-type *SIR4* or *RAD52* genes (Fig. 2a). Similar results were obtained with the *sir2 rad52* and *sir3 rad52* double mutants (Fig. 2b). Because *rad52* mutants are defective in DNA double-strand break repair mediated by homologous recombination¹⁶, these results indicate that the *SIR2*, *SIR3* and *SIR4* genes are involved in DNA double-strand break repair mediated by end-joining.

We have shown that Sir2, Sir3 and Sir4 are involved in illegitimate recombination, end-joining and DNA double-strand break repair, but that Sir1 is not. Sir2, Sir3 and Sir4 form a complex and are involved in silencing at telomeres and *HM* loci, whereas Sir1 only makes a small contribution to silencing at *HM* loci and has no effect at telomeres^{5,6,17–19}. Although the importance of silencing at telomeres is still not fully understood, silencing may protect telomeres from nuclease digestion by making telomeres inactive, as is the case for heterochromatin in higher eukaryotes^{5,6}. DNA ends produced by double-strand breaks may also be protected by a mechanism similar

Table 2 Deletion formation on dicentric plasmid in *sir* mutants

Genotype	Transformation efficiency Leu ⁺ Ura ⁺ per µg (YcPl2)	Number of recombinants Can ^R Cyh ^R Leu ⁺ Ura ⁺ per µg (YdCp2)	Frequency of deletion (s.d.)	Relative increase
Wild type	2.2×10^6	82	$3.6 (1.2) \times 10^{-4}$	1
<i>sir1</i>	3.2×10^6	90	$3.6 (0.8) \times 10^{-4}$	1.0
<i>sir2</i>	3.8×10^6	4.2	$1.2 (0.6) \times 10^{-5}$	0.033
<i>sir3</i>	1.1×10^6	1.2	$8.6 (6.0) \times 10^{-6}$	0.024
<i>sir4</i>	1.0×10^6	0.67	$6.3 (4.1) \times 10^{-6}$	0.018
<i>hdf1</i>	3.4×10^6	2.8	$8.3 (6.7) \times 10^{-6}$	0.023
<i>sir4 hdf1</i>	1.2×10^6	0.83	$6.6 (4.0) \times 10^{-6}$	0.018

The dicentric plasmid YdCp2 and the control monocentric plasmid YcPl2 (0.5 µg each) were introduced into yeast cells by the lithium acetate method. After incubation in YPAD liquid medium for 6 h, the cells transformed with YdCp2 were plated on SD (-Trp -Ura) containing both canavanine and cycloheximide to detect cells with the deleted plasmids, and the cells transformed with YcPl2 were plated on SD (-Trp -Ura) to measure transformation efficiency. Results are an average of six determinations.

to silencing at telomeres, and this process may be essential for end-joining. In telomere silencing, Sir2, Sir3 and Sir4 were found to lack DNA-binding activity, and did not seem to bind directly to DNA^{5,6,20}. The Rap1 protein, which binds to the specific sequences of telomeric regions, is thought to be required for them to bind to telomeres^{5,6,20}. The results of the two-hybrid analysis suggested that Sir4 interacts with Hdf1. As the yeast Ku protein homologue can bind to the ends of double-stranded DNA⁸, Hdf1 may function by recruiting first Sir4, and then Sir2 and Sir3, which interact with Sir4, to the DNA ends produced by the double-strand break.

It is known that the *sir3*, *sir4* and *hdf1* mutants have shortened telomeres^{21,22}. Although a *rap1* mutant has longer telomeres, Rap1 is also likely to be important in telomere maintenance by binding to telomeres^{23,24}. Because Rap1, Sir3 and Sir4 itself interact with the C terminus of Sir4, and we have shown that the C terminus of Sir4 interacts with Hdf1, we envisage that these proteins form a complex in telomeric regions of chromosomes and work together for telomere maintenance.

In addition to these Sir proteins and Hdf1, several other factors are essential for end-joining, including Rad50, Mre11 and Xrs2 (refs 2, 7). However, the relation between the functions of these factors and those of Sir proteins is not yet understood. It is possible that the putative heterochromatin-like structure formed by the Hdf1 and Sir proteins might provide the prerequisite molecular context for the functioning of the Rad50–Mre11–Xrs2 complex in end-joining, but further work is required for the functions of Sir proteins in yeast end-joining to be understood. Our results also suggest that mammalian Sir-related proteins are involved in DNA double-strand break repair and recombination, such as the V(D)J recombination of the immunoglobulin gene. In mammals, transcriptional repression, such as X-chromosome inactivation and genomic imprinting, share some similarities with silencing in yeast⁵. We expect that our understanding of the mechanism of end-joining will be also helped by studies into the functions of Sir protein homologues in mammals. □

Methods

Strains. Recombination assays were performed in yeast strains DH6.61D (*MATa trp1 his3 leu2 ura3 can1 cyh2*; provided by J. W. Szostak) and its derivatives constructed by one-step gene replacement²⁵ of the *SIR1* gene with *sir1::HIS3*, the *SIR2* gene with *sir2::HIS3*, the *SIR3* gene with *sir3Δ::leu2::hisG*, the *SIR4* gene with *sir4::TRP1*, the *HDF1* gene with *hdf1Δ::leu2::hisG* and the *RAD52* with *rad52Δ::hisG*. Yeast strains W303-1A (*MATa ade2-1 trp1-1 his3-11, 15 leu2-3, 112 ura3-1 can1-100*; ref. 26) and its derivatives were constructed as described above and were used to examine sensitivity to ionizing radiation.

Two-hybrid screening. The plasmid pLexHDF1 was constructed using the pBTM116 expression vector containing the *Escherichia coli* *lexA* gene under the control of the *ADHI* promoter and the *TRP1* gene from the 2-µm plasmid (provided by S. Fields). The plasmid generates a fusion protein of the LexA DNA binding domain joined to the full-length *HDF1* gene amplified by PCR. To identify proteins interacting with Hdf1, we used three pools of a library of *S. cerevisiae* genomic DNA fragments fused to the Gal4 activation domain in pGAD plasmids¹³. Yeast strain L40 (*MATa his3Δ200 trp1-901 leu2-3, 112*

ade2 LYS2::(lexAop)₄-HIS3 URA3::(lexAop)₈-lacZ; ref. 27) containing pLexHDF1 was transformed with the yeast library. After 3–4 days of growth on SC (-His-Trp-Leu) plates, His⁺ Trp⁺ Leu⁺ transformants were replica-plated to SC (-His -Trp -Leu +X-gal) plates neutralized by potassium phosphate and incubated for 3 days until blue colonies appeared. Of the 7.9×10^6 transformants, 42 clones were isolated from the library.

Recombination assays. Determinations of the rate of illegitimate recombination on the plasmid YcPl2 and detection of plasmid deletion by end-joining using the dicentric plasmid YdCp2 were performed as described previously^{2,7}. The transformation assay of linearized DNA was performed as described³. Briefly, a yeast strain was transformed by the lithium acetate method with a mixture of 300 ng of linearized pRS316 DNA¹⁴ digested by *Bam*HI and 100 ng of uncut YEpl3 DNA¹⁵ in the presence of 50 µg of denatured salmon sperm DNA as carrier. The transformed cells were diluted and plated onto SC(-Ura) and SC(-Leu). Colonies were counted after incubation at 30 °C for 3 days.

Measurements of sensitivity to ionizing radiation. Sensitivity of yeast strains to ionizing radiation was determined by exposing cell suspensions to a ¹³⁷Cs source (Gammacell 40; Atomic energy of Canada) at 100 rad min⁻¹. Irradiated and non-irradiated cells were diluted in water and spread on YPAD plates to monitor cell survival. Colonies were counted after plates had been incubated at 30 °C for 2–4 days. *SIR4* and *RAD52* plasmids were constructed by cloning PCR-amplified fragments containing the *SIR4* and *RAD52* genes into *Sma*I site of the pRS315 plasmid¹⁴. The experiments were repeated five times.

Received 25 March; accepted 12 June 1997.

1. Tsukamoto, Y., Kato, J. & Ikeda, H. Hdf1, a yeast Ku-protein homologue, is involved in illegitimate recombination, but not in homologous recombination. *Nucleic Acids Res.* **24**, 2067–2072 (1996).
2. Tsukamoto, Y., Kato, J. & Ikeda, H. Budding yeast Rad50, Mre11, Xrs2, and Hdf1, but not Rad52, are involved in formation of deletion mutation on a dicentric plasmid. *Mol. Gen. Genet.* (in the press).
3. Milne, G. T., Jin, S., Shannon, K. B. & Weaver, D. T. Mutations in two Ku homologs define a DNA end-joining repair pathway in *Saccharomyces cerevisiae*. *Mol. Cell Biol.* **16**, 4189–4198 (1996).
4. Boulton, S. J. & Jackson, S. P. *Saccharomyces cerevisiae* Ku70 potentiates illegitimate DNA double-strand break repair and serves as a barrier to error-prone DNA repair pathways. *EMBO J.* **15**, 5093–5103 (1996).
5. Laurenson, P. & Rine, J. Silencers, silencing, and heritable transcriptional states. *Microbiol. Rev.* **56**, 543–560 (1992).
6. Shore, D. In *Telomeres* (eds Blackburn, E. H. & Greider, C. W.) 139–191 (Cold Spring Harbor Laboratory Press, NY, 1995).
7. Tsukamoto, Y., Kato, J. & Ikeda, H. Effects of mutations of *RAD50*, *RAD51*, *RAD52*, and related genes on illegitimate recombination in *Saccharomyces cerevisiae*. *Genetics* **142**, 383–391 (1996).
8. Feldmann, H. & Winnacker, E. L. A putative homologue of the human autoantigen Ku from *Saccharomyces cerevisiae*. *J. Biol. Chem.* **268**, 12895–12900 (1993).
9. Weaver, D. T. What to do at an end: DNA double-strand-break repair. *Trends Genet.* **11**, 388–392 (1995).
10. Boulton, S. J. & Jackson, S. P. Identification of a *Saccharomyces cerevisiae* Ku80 homologue: roles in DNA double strand break rejoining and in telomeric maintenance. *Nucleic Acids Res.* **24**, 4639–4648 (1996).
11. Siede, W., Friedl, A. A., Dianova, I., Eckardt-Schupp, F. & Friedberg, E. C. The *Saccharomyces cerevisiae* Ku autoantigen homologue affects radiosensitivity only in the absence of homologous recombination. *Genetics* **142**, 91–102 (1996).
12. Liang, F., Romanienko, P. J., Weaver, D. T., Jeggo, P. A. & Jasin, M. Chromosomal double-strand break repair in Ku80-deficient cells. *Proc. Natl Acad. Sci. USA* **93**, 8929–8933 (1996).
13. Chien, C. T., Bartel, P. L., Sternglanz, R. & Fields, S. The two-hybrid system: a method to identify and clone genes for proteins that interact with a protein of interest. *Proc. Natl Acad. Sci. USA* **88**, 9578–9582 (1991).
14. Sikorski, R. S. & Hieter, P. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*. *Genetics* **122**, 19–27 (1989).
15. Broach, J. R., Strathern, J. N. & Hicks, J. B. Transformation in yeast: development of a hybrid cloning vector and isolation of the *CAN1* gene. *Gene* **8**, 121–133 (1979).
16. Petes, T. D., Malone, R. E. & Symington, L. S. In *The molecular and Cellular Biology of the Yeast Saccharomyces. Genome Dynamics, Protein Synthesis, and Energetics* Vol. 1 (eds Broach, J. R., Pringle, J. R. & Jones, E. W.) 407–521 (Cold Spring Harbor Laboratory Press, NY, 1991).
17. Pilus, L. & Rine, J. Epigenetic inheritance of transcriptional states in *S. cerevisiae*. *Cell* **59**, 637–647 (1989).

18. Aparicio, O. M., Billington, B. L. & Gottschling, D. E. Modifiers of position effect are shared between telomeric and silent mating-type loci in *S. cerevisiae*. *Cell* **66**, 1279–1287 (1991).
19. Moazed, D. & Johnson, D. A deubiquitinating enzyme interacts with SIR4 and regulates silencing in *S. cerevisiae*. *Cell* **86**, 667–677 (1996).
20. Moretti, P., Freeman, K., Coady, L. & Shore, D. Evidence that a complex of SIR proteins interacts with the silencer and telomere-binding protein RAP1. *Genes Dev.* **8**, 2257–2269 (1994).
21. Palladino, F. et al. SIR3 and SIR4 proteins are required for the positioning and integrity of yeast telomeres. *Cell* **75**, 543–555 (1993).
22. Porter, S. E., Greenwell, P. W., Ritchie, K. B. & Petes, T. D. The DNA-binding protein Hdf1p (a putative Ku homologue) is required for maintaining normal telomere length in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **24**, 582–585 (1996).
23. Sussel, L. & Shore, D. Separation of transcriptional activation and silencing functions of the RAP1-encoded repressor/activator protein 1: isolation of viable mutants affecting both silencing and telomere length. *Proc. Natl Acad. Sci. USA* **88**, 7749–7753 (1991).
24. Buck, S. W. & Shore, D. Action of a RAP1 carboxy-terminal silencing domain reveals an underlying competition between HMR and telomeres in yeast. *Genes Dev.* **9**, 370–384 (1995).
25. Rothstein, R. J. One-step gene disruption in yeast. *Methods Enzymol.* **101**, 202–211 (1983).
26. Thomas, B. J. & Rothstein, R. Elevated recombination rates in transcriptionally active DNA. *Cell* **56**, 619–630 (1989).
27. Hollenberg, S. M., Sternglanz, R., Cheng, P. F. & Weintraub, H. Identification of a new family of tissue-specific basic helix-loop-helix proteins with a two-hybrid system. *Mol. Cell. Biol.* **15**, 3813–3822 (1995).
28. Rose, M. D., Winston, F. & Hieter, P. *Methods in Yeast Genetics, a Laboratory Course Manual* (Cold Spring Harbor Laboratory Press, NY, 1990).
29. Luria, S. E. & Delbruck, M. Mutations of bacteria from virus sensitivity to virus resistance. *Genetics* **28**, 491–511 (1943).
30. Lea, D. E. & Coulson, C. A. The distribution of the numbers of mutants in bacterial populations. *J. Genet.* **49**, 264–285 (1948).

Acknowledgements. We thank S. Fields, S. M. Hollenberg, K. Johzuka, I. Kobayashi, A. Miyajima, J. Rine and J. W. Szostak for providing plasmids and strains, and K. Johzuka for advice on the two-hybrid assay. This work was supported in part by grants to Y.T., J.K. and H.I. from the Ministry of Education, Science, Sports, and Culture of Japan. Y.T. was supported by a postdoctoral fellowship of the Japan Society for the Promotion of Science.

Correspondence and requests for materials should be addressed to H.I. (e-mail: ike@hgc.ims.u-tokyo.ac.jp).

Crystal structure of the breakage–reunion domain of DNA gyrase

João H. Morais Cabral, Andrew P. Jackson, Clare V. Smith, Nita Shikotra, Anthony Maxwell & Robert C. Liddington

Department of Biochemistry, University of Leicester, Leicester LE1 7RH, UK

DNA gyrase is a type II DNA topoisomerase from bacteria that introduces supercoils into DNA^{1,2}. It catalyses the breakage of a DNA duplex (the G segment), the passage of another segment (the T segment) through the break, and then the reunification of the break. This activity involves the opening and closing of a series of molecular ‘gates’ which is coupled to ATP hydrolysis. Here we present the crystal structure of the ‘breakage–reunion’ domain of the gyrase at 2.8 Å resolution. Comparison of the structure of this 59K (relative molecular mass, 59,000) domain with that of a 92K fragment of yeast topoisomerase II (ref. 3) reveals a very different quaternary organization, and we propose that the two structures represent two principal conformations that participate in the enzymatic pathway. The gyrase structure reveals a new dimer contact with a grooved concave surface for binding the G segment and a cluster of conserved charged residues surrounding the active-site tyrosines. It also shows how breakage of the G segment can occur and, together with the topoisomerase II structure, suggests a pathway by which the T segment can be released through the second gate of the enzyme. Mutations that confer resistance to the quinolone antibacterial agents cluster at the new dimer interface, indicating how these drugs might interact with the gyrase–DNA complex.

An alignment of domains in *Escherichia coli* gyrase and yeast topoisomerase II (topo II) is shown in Fig. 1, and their three-dimensional structures are compared in Fig. 2. GyrA59 is the minimal fragment of the A-subunit which, when complexed with

the B-subunit, has DNA-cleavage activity⁴. The carboxy-terminal domain of GyrA is required for the introduction of DNA supercoils, but the GyrA59–GyrB complex supports ATP-dependent relaxation of supercoils in a reaction reminiscent of conventional type II topoisomerases such as yeast topo II (ref. 5).

Each GyrA59 monomer is composed of an N-proximal head and a C-proximal tail. The head contains two domains: one similar to the DNA-binding domain of the catabolite-activator protein (CAP), including the helix–turn–helix (HTH) motif; and a second domain with α/β structure (the ‘tower’ domain) which stands on the CAP-like domain. The tail has a single domain with a helical core. Two long helices (α14 and α18) emanate from this core and connect, together with the C-terminal helix (α19), the head and tail fragments. The head fragment has a very similar structure to yeast topo II (ref. 3). The tail is also structurally conserved at its core, although large surface loops emanating from different points give it a different outward appearance. The tail fragments form the so-called ‘primary’ dimer interface (total buried surface, 2,150 Å²), which is closely conserved in topo II (refs 3, 6), providing a convenient reference point from which to compare the two structures.

The three connecting helices (α14, α18 and α19) adopt very different conformations, leading to large quaternary movements involving a single hinge-point within the helices and rigid body movements of the head fragments (Fig. 2d). Compared with the yeast structure, the angle between α18 and α19 reduces from 120° to 90°. At the same time, the long helix α14 bends by 30° towards the dimer-related molecules and changes the way it packs against α18 and α19, shifting two helical turns on α19. The head fragments, which are rigidly connected to α14, swing towards each other and dock together to form a new contact (the ‘head’ dimer interface), with the active-site tyrosines moving by 26 Å towards and past each other (Fig. 2a, b). The resultant structure is topologically a closed ring, with external dimensions of 100 Å × 40 Å × 100 Å and a central hole of diameter 30 Å. The hole is much smaller than in the topo II structure but it is still large enough to accommodate a DNA duplex.

The ‘head’ dimer interface is dominated by an antiparallel side-by-side packing of the α3 helices (the first helix of the HTH motif) from each monomer, together with their adjoining loops (Fig. 3b, c). The helices pack very closely (mean centre-to-centre distance 7 Å) owing to a preponderance of small side chains. The two abutting surfaces are flat and hydrophobic at the centre, with charged groups at the periphery, burying a total surface area of 1,380 Å². At the top of the interface, the ‘recognition’ helices (α4) make a head-to-head antiparallel dimer contact. Dimer formation creates a large concave surface (100 Å × 30 Å) on the top of the molecule formed by the upper surface of the CAP-like domains and the sides of the two ‘tower’ domains (Fig. 3a). The active-site tyrosines (Tyr 122) are on loops at either end of the dimer interface, 30 Å apart (very similar to their separation in the yeast topo II structure), and sit at the ends of strongly basic grooves created by the dimer-related monomers. Previous biochemical studies have shown that the cleavage of DNA is achieved by a transesterification reaction between the tyrosines and the ‘target’ phosphoryl groups on opposing strands of the DNA backbone, resulting in the tyrosine being covalently attached to the 5′ end of the cleaved segment with a 4-base overhang^{1,2}. The gyrase structure reveals a new cluster of conserved residues (Fig. 3c), juxtaposing Tyr 122 and Arg 121 from one monomer and His 80, Arg 32 and Lys 42 from the other monomer. We propose that this cluster forms the active site of the breakage–reunion reaction, with the other conserved positive charges (Arg 46 and Arg 47) anchoring the non-covalently bound 3′ end of the cleaved DNA.

We modelled a DNA duplex onto the GyrA59 structure by minimizing the distance between the tyrosines and the target phosphates, and by allowing the duplex to bend about a single

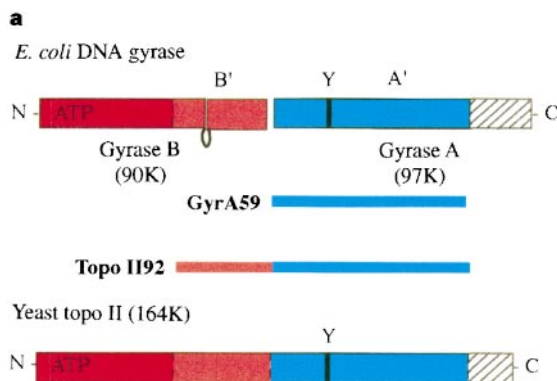


Figure 1 a, Domain organization of *E. coli* gyrase and yeast topoisomerase II. The yeast protein is a homodimer whereas the gyrase is a tetramer with two independent polypeptides. The domains perform similar functions, except for the C-terminal domains, which are not homologous²¹. **b**, Sequence alignment based on the three-dimensional structures. Secondary structure elements are defined by arrows (β -strands) or bars (α -helices), and follow the nomenclature in ref. 3, with additional elements primed. Lower-case letters indicate residues that have not been modelled. Residues that are homologous throughout the known type II topoisomerases⁶ are shaded.

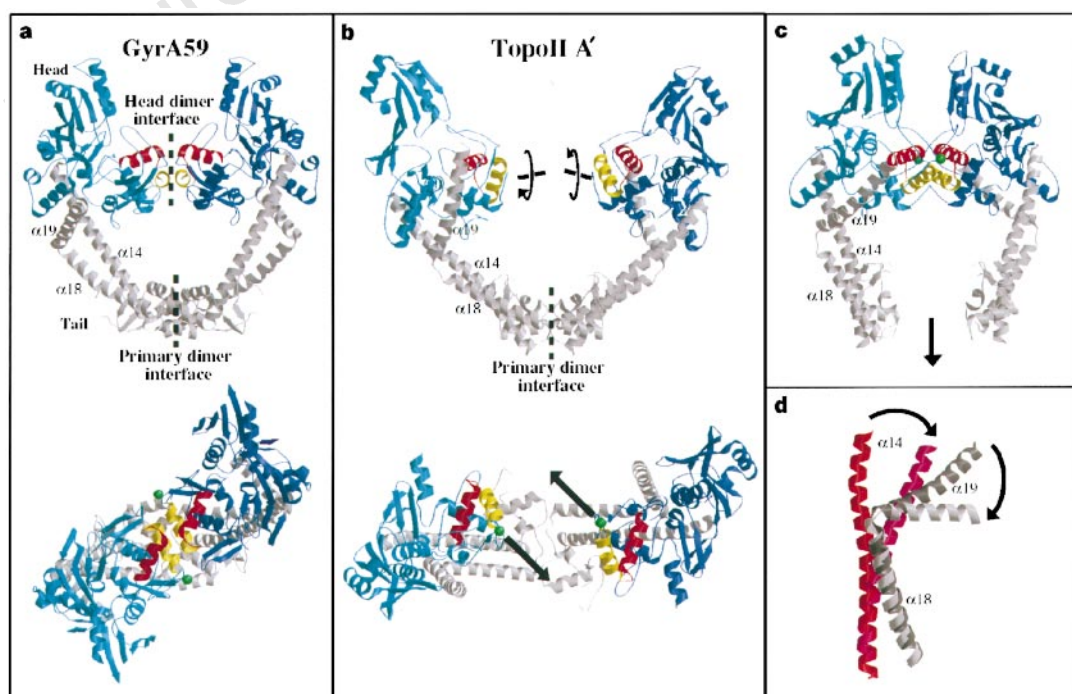
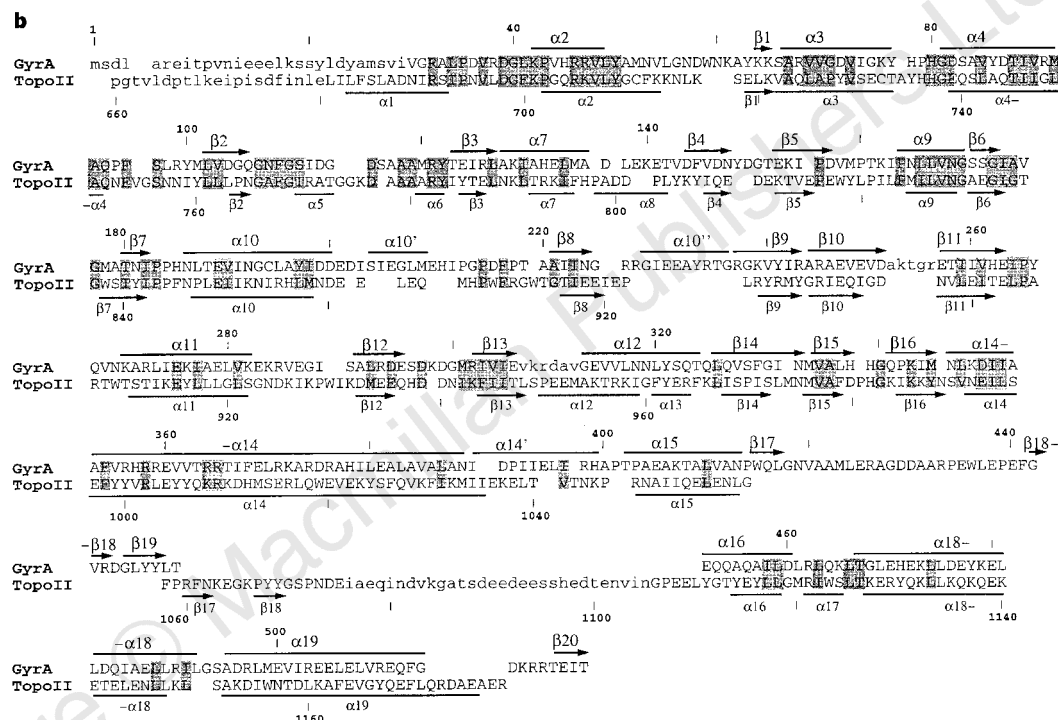


Figure 2 a, b, Orthogonal views of GyrA59 and the yeast topo II A' subunit. Head fragments are shown in blue, helix-turn-helix motifs in yellow (α 3) and red (α 4), tail domains in grey, active-site tyrosines as green spheres. Arrows show the direction of movement from topo II to gyrase. **c**, Opening of the 'second gate' in the primary dimer interface (see text). The A' subunits of yeast topo II were superimposed onto the GyrA59 dimer structure by least-squares fits onto the 'head' fragments. The view is rotated about a vertical axis to show the full extent of the gate (~22 Å). **d**, Two conformations of the connecting helices, superimposed at the 'primary' dimer interface. Arrows show the direction of movement from topo II to gyrase.

axis to follow the concave protein surface. This produced the following model (Fig. 3a): both 'recognition helices' (and most of the quinolone-resistance sites) lie in the central half turn of the DNA major groove; the duplex makes an angle of $\sim 20^\circ$ to the long axis of the dimer face, with the backbone phosphates from opposing strands lying in the basic grooves of the protein. Further protein-DNA contacts are made with the towers and with the $\beta 6$ - $\beta 7$ hairpin, which contains a conserved exposed isoleucine side chain (Ile174) that projects into the minor groove. The DNA curves upwards, towards the presumed location of the B-subunits, consistent with models of gyrase-DNA complexes derived from electron microscopy and neutron scattering^{7,8}. This degree of DNA curvature (radius of curvature ~ 70 Å) has been observed in other DNA-protein complexes⁹.

In our gyrase-DNA model, the target phosphoryl groups are still 7 Å away from the tyrosines, so that either the DNA or the protein must adjust to bridge the gap. A rotation of the gyrase monomers that would bring the tyrosines closer to the phosphates is unlikely as this pushes the 'recognition' helices further apart. A substantial widening of the DNA major groove is a possibility, and this type of distortion has been observed in other DNA-protein complexes^{9,10}. The quinolone drugs bind strongly to the gyrase-DNA complex but only weakly to protein or DNA alone¹¹. The location of the quinolone-resistance mutations at the point where we predict that distortion of the DNA should occur (Fig. 3b, c) is consistent with the idea that the drugs interact with the protein and the DNA bases adjacent to the cleavage site.

The gyrase and yeast topo II structures provide complementary

Table 1 Summary of crystallographic analysis

Data set:	Native	PtCl ₄	PtCl ₄	PIP	PIP	MMN
Wavelength (Å):	1.54	1.54	1.067	1.54	0.87	1.54
Resolution (Å)	2.8	3.25	3.5	3.0	2.8	4.5
Completeness (%)	100	98	92	100	84	97
Redundancy	7	2	2	3	4	2
<i>R</i> _{sym} (%)†	4.7	7.4	12.6	7.5	7.9	7.3
No. of sites		3	2	5	6	2
<i>R</i> _{iso} (%)‡		17.6	15.4	11.5	14.0	10.9
<i>R</i> _{cullis} (%)§		72	78	68	65	86
Phasing power¶		1.6	1.3	1.8	1.9	0.9
Overall figure of merit:	0.58					
Refinement statistics:						
Resolution (Å)		<i>R</i> factor (%)	<i>R</i> _{free} (%)	r.m.s. bond lengths (Å)	r.m.s. bond angles	No. of atoms
15-2.8	22.6	31.0	0.006	113	3,774 + 15 H ₂ O	No. of reflections
						16,367

PIP, di-μ-iodobis(ethylenediamine) diplatium(II) nitrate; MMN, methylmercuric nitrate; PtCl₄, potassium tetrachloroplatinate (II).

**R*_{sym} = $\sum \sum |I_j - \langle I \rangle| / \sum I_j$.

†*R*_{iso} = $\sum |F_{ph} - F_p| / \sum F_p$.

‡*R*_{cullis} = $\sum |F_{ph} \pm F_p| - |F_{hc}| / \sum |F_{ph} \pm F_p|$.

§ Phasing power = $\langle F_{ph} \rangle / E$.

||*R*_{factor} = $\sum |F_{ph} - F_p| / \sum F_p$; *R*_{free} is the same as *R* factor but calculated on the 10% of reflections excluded from refinement; *I*_j is the measured intensity for reflection *j*, $\langle I \rangle$ is the mean intensity, *F*_p is the structure factor for the native, *F*_{ph} is the structure factor for derivative, *F*_{hc} is the calculated structure factor for the heavy atom, $\langle F_p \rangle$ is the root-mean-square heavy-atom structure factor, *E* is the residual lack of closure.

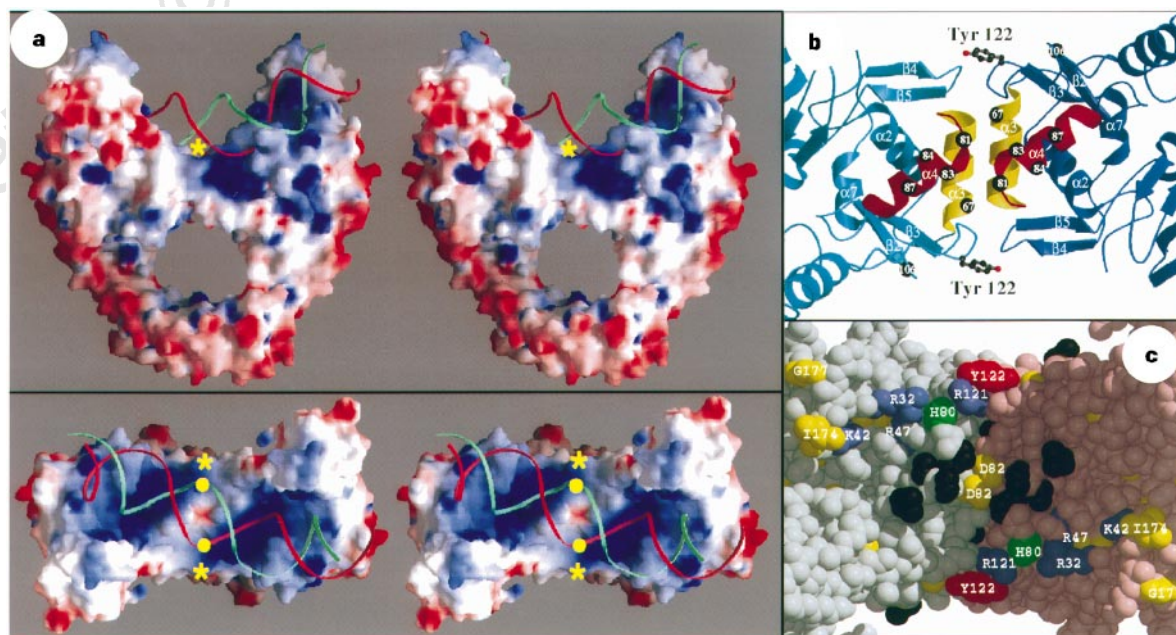


Figure 3 **a**, GRASP²² electrostatic surface potential of the GyrA59 dimer (orthogonal stereo views): negatively charged surfaces are in red, positively charged surfaces in blue. The DNA backbone is shown as a green and red ribbon; active-site tyrosines as yellow stars, target phosphoryl groups as yellow dots. **b**, Close-up of the 'head' dimer interface (colour scheme as in Fig. 2), with some secondary

structure elements indicated. Quinolone-resistance sites are shown as numbered black spheres. **c**, Space-filling model, viewed as in **b**. Monomers are shown in pale green and pink. Quinolone-resistance sites are in black. Residues picked out in other colours are highly conserved within the topoisomerase family.

segment. □

Methods

guide) and phase recombination.

and RENDER²⁰.

Received 20 February; accepted 5 June 1997.

12. Roca, J. & Wang, J. C. DNA transport by a type II DNA topoisomerase: evidence in favor of a two-gate mechanism. *Cell* **77**, 609–616 (1994).

22. Nicholls, A. GRASP: Graphical representation and analysis of surface properties. (Columbia University, New York, 1992).

This work was supported by the BBSRC. A.M. is a Lister Institute Jenner Fellow.

Correspondence and requests for materials should be addressed to R.C.L. (e-mail: rcl6@le.ac.uk). Coordinates have been deposited with the Protein Data Bank (1AB4).

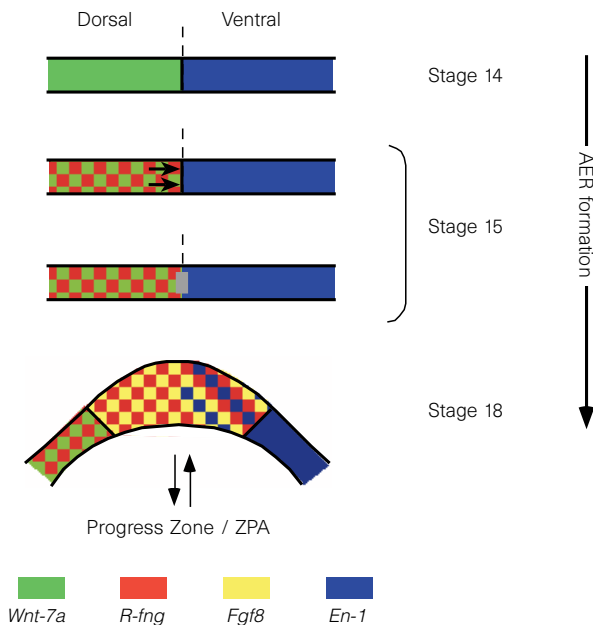
erratum

***Radical fringe* positions the apical ectodermal ridge at the dorsoventral boundary of the vertebrate limb**

**Concepción Rodríguez-Esteban, John W. R. Schwabe,
Jennifer De La Peña, Bryon Foys, Brian Eshelman
& Juan Carlos Izpisua Belmonte**

Nature **386**, 360–366 (1997)

In Fig. 7 of this Article, the two arrows indicating signalling by *R-fng* at stage 15 were unclear; the correct figure is shown here. □



DNA technology

This round-up of molecular tools includes a thermal printer module, a PCR-product amplification detection system, plasmid purification kits, and a system for *in situ* hybridization, RT-PCR, PRINS and immunohistochemistry.

Millennium markers

From Ambion

Formulated to give accurate size determinations of single-stranded RNA molecules in northern blot analysis

This is a mixture of *in vitro* transcribed RNA — the transcripts include lengths of 0.5, 1, 1.5, 2, 2.5, 3, 4, 5, 6 and 9 kb. These markers may be stained with ethidium bromide during or after electrophoresis. After transfer to a membrane, the markers may be visualized with methylene blue or they may be probed with labelled λ sequence to generate a signal on film. The large number and even spacing of the Millennium Marker bands should allow the construction of accurate standard curves for determining the size of unknown mRNA species. Millennium Markers are also available in a biotinylated format for non-isotopic detection.

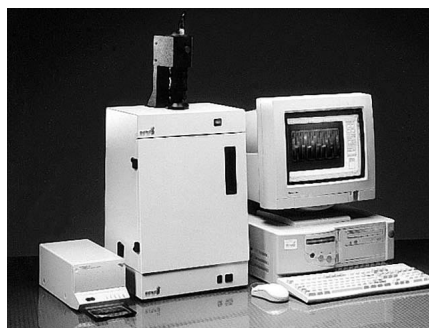
Reader Enquiry No. 100

UVIpro

From UVItec

A gel analysis and documentation system

This system acts as a fully integrated system for image acquisition, filing, printing and analysis, yet it retains a modular design that allows flexibility in specification and price. At the heart of the UVIpro system is a compact CCD camera, which is neatly interfaced to a Pentium PC via a single cable and a PCI acquisition card. From the user friendly image-acquisition software, the user controls parameters such as camera gain, offset and integration time, allowing optimization of the video image in 'real time'. For detection of very faint fluorescent bands, integration times of up to two minutes are possible. Once captured, the image may be enhanced by fine adjustment of brightness and contrast and may be annotated with text and symbols. The UVIphoto software also includes features such as good laboratory practice facilities, the ability to



The UVIpro gel analysis and documentation system from UVItec.

save and recall acquisition parameters and the option of automatic addition of the current date and time to an image. Captured images may be saved in a variety of file formats (for example, TIFF, JPEG and GIF) or printed directly to an analog thermal printer of which two models are available. These include a standard-quality thermal printer and one that produces superior, near-photographic quality images on plastic film.

Reader Enquiry No. 101

EPS 1000

From Hoefer Pharmacia Biotech

A versatile power supply for general applications

Pharmacia Biotech has introduced the EPS 1000 power supply, which has been designed for a wide range of polyacrylamide gel electrophoresis applications (mini and standard vertical slab gels, or ultrathin flatbed gels for Multiphor or GenePhor). The unit delivers 5–1,000 V, 1–400 mA, or 1–100 W at constant current, constant voltage, or constant power with automatic crossover. There is parameter increment resolution of 1 mA, 1 V and 1 W, and three timer options that provide precise settings for reproducible results. The voltage, current and elapsed time parameters are displayed simultaneously on a backlit 2 × 16-character LCD display. Two sets of 2-mm recessed power outlets are available for running two electrophoresis units in parallel. The instrument is compact (9.5 cm tall × 25 cm wide × 38.5 cm deep) and is stackable.

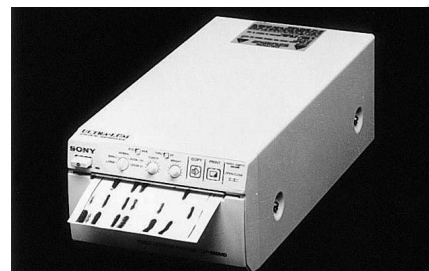
Reader Enquiry No. 102

Thermal printer

From Ultra-Lum

A 256-grayscale thermal printer that is said to be easy to use and saves on bench space, time and money

This printer produces publication-quality prints in about 3.9 s at a fraction of the cost when compared with film-based systems (a stated 10 cents (US) an image). This printer provides high resolution (8.9 dots mm⁻¹), ×1.5 and ×2 zoom capability, and portrait/landscape image orientation. The contrast and brightness dials can be set to control both the printer and video display simultaneously — so that what is seen is exactly what is printed. The thermal printer also complements Ultra-Lum's modular line of gel documentation systems. With the modular approach, multi-user laboratories can print and digitally document colour and



UltraLum's new high-resolution thermal printer.

fluorescent stains, autorads, blots and ELISAs. The printer automatically adjusts to the supplied voltage (115, 100, 220 or 240 V).

Reader Enquiry No. 103

Sunrise

From Oncor

A PCR product amplification detection system that permits the simultaneous amplification and detection of nucleic acids within the PCR tube

The method is based on the incorporation of energy-transfer-labelled hairpin 'primers' into the amplified DNA product. The Sunrise UniPrimer I is available for universal amplification of any target. It consists of a hairpin primer designed in such a way that a fluorescent signal is generated only when the primer is unfolded by incorporation into the amplification product. Unincorporated Sunrise UniPrimer I has virtually no fluorescent signal, eliminating the need to purify the PCR product before detection. The 'closed-tube' format of the Sunrise system should help eliminate PCR contamination, post-PCR sample manipulation (such as gel analysis) and should help accommodate high-throughput analysis of amplified material.

Reader Enquiry No. 104

CycleSeal

From Robbins Scientific

An easy and convenient way to seal PCR plates for thermal cycling

CycleSeal is a single pre-cut, silicon-like sheet with a removable plastic backing. The heat and pressure from the thermal cycler lid moulds the sealant around the well openings, providing an effective evaporative seal with no well-to-well contamination. It can be used on any 96-well thermal cycler, with a heated lid, including those from Perkin-Elmer and MJ Research, and is compatible with all Robbins' brands of CyclePlates, including the new CyclePlate-384 and CyclePlate-192 products.

Reader Enquiry No. 105

InViSorb

From InViTek

A new generation of spin column kits for plasmid purification

The new kits combine the company's batch method with the convenience, speed and high throughput of spin columns. There are said to be no toxic reagents involved, such as phenol, chloroform, CsCl or ethidium bromide. Depending on the plasmid copy number and the required amount of plasmid DNA, the manufacturer offers a choice between mini, midi and maxi kits. Each mini kit contains either 25 or 50 columns for small-scale preparations (700- μ l capacity). The midi and maxi kits have 3.5-ml and 15-ml capacities, respectively.

Reader Enquiry No. 106

Cleanaprep

From ICN

Two products that offer plasmid DNA purification (Cleanaprep) and DNA isolation (Cleanaprep II)

Cleanaprep has been formulated to be a rapid, inexpensive and simple way of producing high yields (up to 20 μ g) of purified plasmid DNA from as little as 2 ml of culture. The kit is safe to use as it contains no harmful or toxic reagents, such as phenol or phenol/chloroform. The spin filter contained in the kit binds plasmid DNA, whereas unbound genomic DNA and protein are removed by microcentrifugation and the bound plasmid DNA can then be eluted from the filter. The purified DNA is immediately ready for use in sequencing or transformations; it can also be ligated by restriction enzymes, when required. The kit is said to perform well for both high- and low-copy plasmids. Cleanaprep kits are available in 25, 100 or 250 preparation size packs. Cleanaprep II can be used for isolation of up to 15 μ g of plasmid, linear or any other DNA from contaminants such as proteins, oligonucleotides, RNA and double-stranded DNA (of less than 45 bp), nucleotides and primers.

Reader Enquiry No. 107

In situ system

From Hybaid

A system created for in situ hybridization, in situ reverse-transcription PCR, PRINS and immunohistochemistry

The system is based around a slide rack, which allows the simultaneous manipulation of 20 slides through the whole experimental protocol. Hybridization and thermal cycling are conducted on the OmniSlide thermal cycler and the subsequent wash steps in the wash module. For the scientist who does not want to make the commitment to the large-scale system, the company also supplies flat



Hybaid's *in situ* system is also available for small-scale operations using Touchdown.

blocks for the Touchdown thermal cycler, which has space for four slides.

Reader Enquiry No. 108

Thermo-Fast

From Advanced Biotechnologies

A 384-well plate with a well capacity of 40 μ l and a working volume of 25 μ l

The plates are made of virgin polypropylene and are skirted for compatibility with automated systems, providing an even surface for robotic arms to clamp onto and sufficient space for each plate to be labelled or barcoded for identification. In addition, the skirt has eight strategically placed holes to facilitate plate positioning and removal from the thermal cycler block. Though the plate is rigid, the wells have thin walls for rapid heat transfer in the thermal cycler. The wells have a raised rim to ensure contact is made with the sealing medium, avoiding problems associated with evaporation, a key consideration when dealing with such small volumes. It is possible to seal Thermo-Fast 384 with a variety of sealing methods, including adhesive seals, single piece mat devices. Other leading thermal cycler manufacturers, such as Hybaid, Techne and Stratagene, are also developing heating blocks for their new machines, which will accommodate the 384-well plate.

Reader Enquiry No. 109

Plasmid prep scale up

From Nucleon Biosciences

A kit for large-scale plasmid preparations suitable for anywhere between 30–500-ml overnight cultures

The kit makes use of a resin that partitions away the contaminating cell lysate, leaving pure DNA in aqueous solution, which is then precipitated in ethanol. Variable yields and overloading have been reported to be the problems associated with fixed-sized cartridges/columns that bind the DNA directly. This system compensates for fluctuations in cell culture densities and routinely provides yields of 2 mg from a 100-ml culture of high copy number plas-

mid and up to 16 mg from a 500-ml culture. According to the manufacturer, the kit will also give consistently high yields with higher molecular weight plasmids.

Reader Enquiry No. 110

HTS 7000

From Perkin-Elmer

A rapid, high-throughput bioassay reader for screening applications

The new reader has been designed to be a versatile, optimized system for rapid, high-throughput screening of a large array of small-volume samples, which are routinely performed in many molecular, cell biology and biochemistry laboratories. The HTS 7000 reader complements other industry leading systems from Perkin-Elmer, such as those for PCR and automated DNA sequencing and synthesis. Applications for the new HTS 7000 include DNA and protein quantification; drug screening and therapeutic monitoring; immunoassays; cell function studies; and DNA hybridization, gene expression, recombinant enzyme and reporter gene assays. The HTS 7000 includes applications software designed for processing and tracking up to hundreds or thousands of samples per day. Top or bottom reading of luminescence or through-well absorbance measurements is achieved automatically through simple software selection. Users with few samples can take advantage of the convenience of multichannel pipetting in strip wells and perform traditional cuvette-based analyses much more rapidly with the HTS 7000. Stored plate maps contain sample identification information and all the necessary details for use of standards, controls and replicates.

Reader Enquiry No. 111

These notes are compiled by Brendan Horton from information provided by the manufacturers. For more details, fill in the reader service card bound inside the journal.

ADVERTISEMENT

2-Hybrid System Media

Over 400 recipes and formulae in stock

CONVENIENCE

CONSISTENCY

ECONOMY

RELIABILITY

Bio 101

Contact Bio 101 or your local distributor for ordering & information
 (800) 424-6101 • fax: (760) 598-0116
 web: www.bio101.com • e-mail: info@bio101.com

READER ENQUIRY NO. 12